

VISIT...

LANZAROTE
Caliente.COM

В. С. МИХАЙЛОВ

ТЕОРИЯ УПРАВЛЕНИЯ

*Допущено
Министерством высшего
и среднего специального
образования УССР
в качестве учебного пособия
для студентов вузов,
обучающихся по специальности
«Системы автоматизированного
проектирования»*

Киев

Головное издательство
издательского объединения
«Вища школа»

1988

ББК 32.965я73

М69

УДК 658.52.011.56 (07)

Рецензенты: доктор технических наук профессор Ю. П. Зайченко (Киевский политехнический институт), доктор технических наук профессор А. А. Волков (Киевский институт инженеров гражданской авиации)

Редакция литературы по строительству и архитектуре
Редактор В. В. Тронько

Михайлов В. С.

М69 Теория управления.—К.: Выща шк. Головное изд-во, 1988. — 312 с.; 26 табл., 80 ил. — Библиогр.: 35 назв.

ISBN 5—11—001791—3

Даны основы теории управления техническими объектами и производственно-экономическими системами. Рассмотрены принципы построения автоматических и автоматизированных систем, принципы и законы управления, математическое описание линейных и нелинейных динамических систем, методы исследования устойчивости и качества управления. Особое внимание уделено основам оптимального управления, использованию ЭВМ в теории и практике управления, математическому программированию как совокупности методов решения задач управления в сложных системах. Приведены примеры построения гибких автоматизированных производств и систем автоматизированного проектирования.

Для студентов вузов, обучающихся по специальности «Системы автоматизированного проектирования».

0605010201—187

М ————— БЗ—45—4—87
М211(04)—88

ББК 32.965я73

ISBN 5—11—001791—3

© Издательское объединение
«Выща школа», 1988

Возрастающие объемы общественного производства непрерывно усложняют задачи управления разнообразными объектами и процессами. Проблемы управления особенно усложняются в системе «человек — машина», где наряду с задачами управления техническими объектами и технологическими процессами остро ставятся задачи эффективного управления системами организационно-технического, экономического и социального характера. Поэтому в вузах для соответствующих специальностей необходима учебная дисциплина, формирующая общую базу для решения задач управления в технических и технико-экономических системах.

В учебных планах подготовки инженеров по специальности 22.03 «Системы автоматизированного проектирования» (САПР) в качестве указанной учебной дисциплины предусмотрен курс «Теория управления». Он относится к числу важнейших специальных учебных дисциплин, в значительной мере определяющих профессиональный уровень инженера-системотехника — специалиста по САПР.

В предлагаемом учебном пособии рассматриваются основные определения и понятия, необходимые для усвоения принципов разработки и функционирования систем управления, методы математического описания элементов и систем, исследования режимов работы систем управления, выбора управляющих решений в технических и технико-экономических системах, в том числе в условиях неопределенности. Особое внимание уделяется задачам оптимального выбора управляющих решений при наличии множества вариантов решений.

Мировоззренческая направленность курса обеспечивается использованием основ диалектической теории познания, ориентацией на принципы единства материального мира путем широкого использования аналогичного математического описания объектов и процессов технического, экономического и биологического характера. Основой для написания книги послужил многолетний опыт научной, преподавательской и производственной работы автора в области систем автоматического и автоматизированного управления, в частности, накопленный при чтении курсов «Теория автоматического управления», «Основы построения АСУ», «Основы научных исследований», «Вычислительная техника и программирование», «Введение в специальности АСУ и САПР».

При работе над учебным пособием использована обширная отечественная и зарубежная литература по теории автоматического управления, исследованию операций, технической кибернетике, основам автоматизированных систем управления и систем автоматизированного проектирования. Список литературы приведен в конце книги.

Все замечания по совершенствованию учебного пособия просьба направлять по адресу: 252054, Киев-54, ул. Гоголевская, 7, издательство «Выща школа».

Определим понятие «управление» как процесс приведения заданных объектов в состояние, отвечающее поставленным целям. При этом объект, которым нужно управлять, назовем *объектом управления*.

Повседневная деятельность человека — производственная, общественная, бытовая, административная — связана с управлением объектами разнообразного характера. Управлять можно транспортным средством, металлорежущим станком, строительным трестом, школой, предметом домашнего обихода и т. п. Из выбранных примеров объектов управления видно, что понятие «управление» чрезвычайно широкое и многообразное.

Действительно, задача управления, например, домашним пылесосом существенно отличается от задачи управления кораблем в море, а управление домашним хозяйством нельзя сравнивать с управлением, например, строительным трестом. Несмотря на это, любое управление имеет общие черты, к числу которых относятся: наличие информации о задачах управления — цель управления; наличие информации о результатах управления, т. е. о реакции объекта управления на управляющие воздействия; анализ информации о состоянии объекта и степени его приближения к требуемому состоянию и выработка необходимых управляющих решений; реализация управляющих воздействий.

Наука об общих закономерностях процессов управления составляет суть теории управления. Теория управления тесно связана с понятием «кибернетика» — наукой об управлении, связи и переработке информации. Именно теоретическая (общая) кибернетика рассматривает вопросы управления абстрактно, безотносительно к реальной физической природе объектов управления, находит общие методы подхода к изучению систем, существенно отличающихся по своей природе: технических, биологических, социальных.

Задачи управления и соответствующие им способы и средства решения появились вместе с возникновением цивилизации. Есть основание считать, что термин «кибернетика» как управление в общем смысле применял древнегреческий философ Платон (428—348 гг. до н. э.). Действительно, греческое слово *kybernētika* переводится как «искусство управления».

Длительное время теория и практика управления развивались под влиянием необходимости решения задач управления техниче-

скими объектами. Эти задачи стали достаточно актуальными после появления первых промышленных регуляторов И. Ползунова (1765 г.) и Дж. Уатта (1782 г.). Дальнейшее развитие техники привело к возникновению одного из направлений теории управления — теории автоматического управления (регулирования). Во второй половине XIX в. появился ряд крупных научных работ по теории автоматического регулирования, не потерявших своего значения до настоящего времени. К числу таких работ следует отнести статью русского математика, акад. П. Л. Чебышева «О центробежном уравниателе» (1871 г.). Основы современной линейной теории автоматического регулирования были заложены в трудах крупного русского ученого И. А. Вышнеградского «О регуляторах прямого действия» (1873 г.), «Об общей теории регулирования» (1876 г.) и др. Важной заслугой И. А. Вышнеградского является правильная постановка проблемы управления машинной системой, при которой объекты управления — машина и регулятор рассматриваются как единая динамическая система. Он разработал эффективные методы анализа устойчивости системы регулирования, позволяющие проследить влияние на устойчивость различных конструктивных параметров системы регулирования.

В конце XIX и начале XX века получают значительное развитие многие отрасли промышленности, транспорт, системы связи. Появляются все более сложные системы автоматики, что потребовало дальнейшего развития теории автоматического управления. Этот период отмечен выдающимися работами русского ученого А. М. Ляпунова (1857—1918 гг.) в области устойчивости динамических систем. Большое влияние на развитие теории управления оказали также работы Н. Е. Жуковского (1847—1921 гг.), среди которых следует отметить его труд «Теория регулирования хода машин», опубликованный в 1909 г.

В первые годы Советской власти одним из первых научных трудов по теории автоматического управления стала работа И. Н. Вознесенского (1887—1946 гг.) «О регуляторах непрямого действия». Впоследствии член-корреспондент АН СССР И. Н. Вознесенский стал основателем одной из школ по теории автоматического управления.

В годы первых пятилеток наша страна переживала бурное промышленное развитие. В этот период создавались новые и переводились на современную базу старые отрасли промышленности, что требовало интенсивного развития теории управления. Из теоретических работ этого периода заметную роль сыграла опубликованная в 1938 г. работа А. В. Михайлова «Метод гармонического анализа в теории регулирования», положившая начало широкому применению частотных методов исследования систем автоматического управления.

В период Великой Отечественной войны все усилия советских ученых и инженеров были направлены на создание систем вооружения и систем, обеспечивающих выпуск вооружения. Высокий уровень автоматики и автоматизации производственных процессов в значительной мере способствовал разгрому врага.

Большую роль теория и практика автоматического управления сыграла в деле восстановления народного хозяйства в послевоенный период. Было разработано и внедрено в практику управления множество новых элементов, таких как электромашинные и магнитные усилители, разнообразные электронные, механические, гидравлические, пневматические и другие устройства. Начиная с 50-х годов в практику управления все шире стали внедряться электронно-вычислительные машины (ЭВМ), сначала аналоговые, а затем — цифровые, что подняло управление на новую, качественную ступень.

На развитие теории управления в послевоенный период существенное влияние оказали работы Н. Винера и К. Шеннона, сыгравшие важную роль в формировании кибернетики как науки, касающейся всей системы человеческих знаний.

В конце 50-х и начале 60-х годов начинается новый этап развития производительных сил, получивший название научно-технической революции (НТР). Ее можно определить как коренное, качественное преобразование производительных сил на основе превращения науки в ведущий фактор развития общественного производства. Важнейшими предпосылками возникновения НТР явились: решение проблемы расщепления атомного ядра и создание источников атомной энергии — практически неисчерпаемых источников энергии; создание ЭВМ; крупные успехи радиоэлектроники, кибернетики, космической техники.

НТР изменила весь облик общественного производства, условия, характер и содержание труда, структуру производительных сил, отраслевую и профессиональную структуру общества. Она оказала и продолжает оказывать огромное воздействие на все стороны деятельности человека — производство, быт, культуру, психологию, взаимодействие с природой. Огромные, принципиальные изменения внесла НТР в теорию и практику управления общественным производством, что потребовало принципиально новых методов и средств управления.

Среди основных факторов, обусловивших коренные изменения в проблемах управления, отметим стремительный рост номенклатуры промышленных изделий, сопровождающийся значительным усложнением изделий и технологии их производства, а также частую сменяемость изделий. Характерно, что моральное старение изделий теперь наступает значительно раньше их физического износа, что требует существенной перестройки технологии, уменьшения сроков разработки и внедрения промышленных изделий. Наряду с отмеченными факторами, на все стороны общественного производства значительно влияет сокращение сырьевых ресурсов, обострились проблемы охраны окружающей среды. Все это существенно усложнило проблемы управления народным хозяйством.

Крупные масштабы производства потребовали переработки огромных объемов информации, вовлекая в сферу управления большие массы людей. Информационный взрыв заставил коренным образом изменить методы и средства управления как в экономических, так и в технических системах. Возникло и стало бурно

развиваться новое направление теории управления — управление большими (сложными) системами. Теоретической основой этого управления стали экономико-математические методы, а технической — электронно-вычислительные машины. Появились человеко-машинные системы — автоматизированные системы управления (АСУ), автоматизированные системы управления технологическими процессами (АСУТП), системы автоматизированного проектирования (САПР). Основными теоретическими методами решения задач управления в сложных системах стали методы математического программирования, реализуемые с помощью ЭВМ. Таким образом, ЭВМ стала не только одним из основных элементов систем управления, но и средством решения задач управления.

Развитие сложных (больших) систем обусловило стремительное развитие и широкое внедрение в практику одного из важнейших направлений современной теории управления — экономико-математических методов, использующих математические средства для описания экономических объектов и процессов. Начало этому направлению было положено работами советского математика и экономиста, академика АН СССР Л. В. Канторовича. В 1939—40 гг. Л. В. Канторович положил начало линейному программированию — теории и методам решения экстремальных задач с ограничениями. Его научный труд «Математические методы организации и планирования производства» сыграл важную роль в создании теории оптимального управления народным хозяйством. Оптимальное управление обеспечивает наилучшее (оптимальное) с некоторой точки зрения функционирование управляемого объекта, существенно повышая технико-экономические показатели систем управления.

Огромное влияние на развитие теории оптимального управления оказали работы академика АН СССР Л. С. Понтрягина и его сотрудников В. Г. Болтянского, Р. В. Гамкрелидзе, Е. Ф. Мищенко. Сформулированный ими принцип максимума Понтрягина известен во всем мире и широко используется в теории оптимальных процессов. Основные результаты теории оптимального управления содержатся в монографии указанных авторов «Математическая теория оптимальных процессов» (1961 г.), получившей мировую известность.

Развитию теории оптимального управления в значительной степени способствовали идеи метода динамического программирования, основы которого разработаны американским ученым Р. Беллманом и его сотрудниками.

Роль теории и практики управления особенно возрастает в настоящий период, характеризующийся требованиями всемерного ускорения научно-технического прогресса. Стратегический курс КПСС на ускорение социально-экономического развития страны требует коренного технического перевооружения производства на базе современных средств автоматизации, прежде всего электроники и вычислительной техники. Определяя понятие «ускорение», Генеральный секретарь ЦК КПСС товарищ М. С. Горбачев подчеркнул: «Суть его — в новом качестве роста: всемерной интенсификации производства на основе научно-технического прогресса, структур-

ной перестройки экономики, эффективных форм управления...» (Политический доклад Центрального Комитета КПСС XXVII съезду Коммунистической партии Советского Союза).

В Основных направлениях экономического и социального развития СССР на 1986—1990 годы и на период до 2000 года подчеркивается: «Шире исследовать проблемы механики, вопросы теории автоматизации производства... Широко внедрять гибкие переналаживаемые производства и системы автоматизированного проектирования, автоматические линии, машины и оборудование со встроенными средствами микропроцессорной техники, многооперационные станки с числовым программным управлением, робототехнические, роторно-конвейерные комплексы».

Теория управления может и должна внести заметный вклад в реализацию стратегического курса КПСС на ускорение социально-экономического развития нашей страны.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ПОНЯТИЯ ТЕОРИИ УПРАВЛЕНИЯ

1.1. ОБЪЕКТЫ И ПРОЦЕССЫ УПРАВЛЕНИЯ. ВИДЫ УПРАВЛЕНИЯ

Характеризующую состояние объекта управления ОУ (рис. 1.1) величину $X_{\text{вых}}$, которой нужно управлять, назовем *выходной*, или *управляемой (регулируемой), величиной*. В общем случае выходная величина представляет собой вектор $X_{\text{вых}} = \{x_{\text{вых}1}, x_{\text{вых}2}, \dots, x_{\text{вых}n}\}$. Всякий реальный объект находится под влиянием окружающей среды, которое проявляется в *возмущающих воздействиях* $F = \{f_1, f_2, \dots, f_k\}$, приложенных к ОУ. Эти воздействия стремятся вывести ОУ из заданного состояния, для приведения в которое к объекту необходимо прилагать управляющие воздействия, которые назовем *входными воздействиями* $X_{\text{вх}} = \{x_{\text{вх}1}, x_{\text{вх}2}, \dots, x_{\text{вх}m}\}$. Место приложения входного воздействия назовем *входом*.

Таким образом, задача управления объектом сводится к тому, чтобы выбрать необходимые входные воздействия $X_{\text{вх}}$, которые при любых возмущающих воздействиях F обеспечивали бы заданное значение выходной величины $X_{\text{вых}}$. В некоторых случаях задача управления объектом сводится к задаче регулирования — обеспечению постоянства выходной величины.

Для простоты будем рассматривать объекты управления, характеризующиеся одной выходной величиной $x_{\text{вых}}$, одной входной $x_{\text{вх}}$ и локализованным возмущающим воздействием f .

Чаще всего управляющее воздействие не может быть приложено непосредственно к объекту управления, а формируется с помощью технических устройств, называемых *регуляторами*. Регулятор Р и объект управления ОУ составляют систему управления (регулирования) (рис. 1.2, а).

Возможно управление по разомкнутому и по замкнутому циклу.

В первом случае между выходной и входной величинами отсутствует непосредственная связь. Она осуществляется через регулятор и объект управления. Управление по разомкнутому циклу обычно выполняется с помощью оператора Оп (рис. 1.2, а), который, наблюдая за значением выходной величины $x_{\text{вых}}$, устанавливает необходимые значения входной величины $x_{\text{вх}}$. Управление по разомкнутому циклу называют иначе управлением с разомкнутой обратной связью. Такое управление с непременным участием человека (оператора) называется *ручным управлением*.

При управлении по замкнутому циклу изменения выходной величины передаются на вход системы с помощью совокупности устройств, называемых *обратной связью* (ОС) (рис. 1.2, б). Вход-

ной величиной обратной связи является выходная величина объекта управления $x_{\text{вых}}$. Выходную величину обратной связи обозначим $x_{0.c}$. Эта величина, накладываясь на входную величину, формирует регулирующее воздействие x_p на входе регулятора, который, в свою очередь, формирует управляющее воздействие $x_{0.y}$, прикладываемое к объекту управления. Образованная таким образом обратная связь называется *главной обратной связью*.

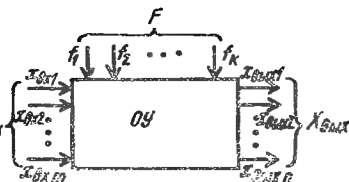


Рис. 1.1. Объект управления

Для осуществления управления по схеме рис. 1.2, б необходимо обеспечить условие

$$x_p = x_{\text{вх}} - x_{0.c}, \quad (1.1)$$

т. е. выходная величина обратной связи должна вычитаться из величины на входе системы. Такая обратная связь называется *отрицательной*. Вычитание управляющих воздействий производится в суммирующем устройстве Σ (рис. 1.2, б).

Рассмотрим подробнее управление по замкнутому циклу. Пусть задача управления сводится к поддержанию величины $x_{\text{вых}}$ на заданном уровне. Подбором соответствующего значения величины $x_{\text{вх}}$ устанавливается требуемое значение величины $x_{\text{вых}}$. Далее процесс регулирования $x_{\text{вых}}$ осуществляется без участия оператора (человека), т. е. автоматически. Действительно, если $x_{\text{вых}}$, например, превысит заданные значения, то на вход ОС поступит управляющий сигнал, больший того, какой имел место при заданном значении выходной величины $x_{\text{вых}}$. Это приведет к уменьшению величин x_p и $x_{0.c}$ [см. уравнение (1.1)], что повлечет за собой уменьшение величины $x_{\text{вых}}$. Рассуждая аналогично, можно убедиться в том, что при уменьшении $x_{\text{вых}}$ под влиянием возмущения f процесс управления также будет направлен на стабилизацию $x_{\text{вых}}$.

Управление по замкнутому циклу называют также *управлением с замкнутой отрицательной обратной связью*.

Если выходная величина обратной связи суммируется с величиной $x_{\text{вх}}$, то такая обратная связь называется *положительной*. При этом

$$x_p = x_{\text{вх}} + x_{0.c}. \quad (1.2)$$

Если рассмотреть процесс управления по схеме рис. 1.2, б, основны-

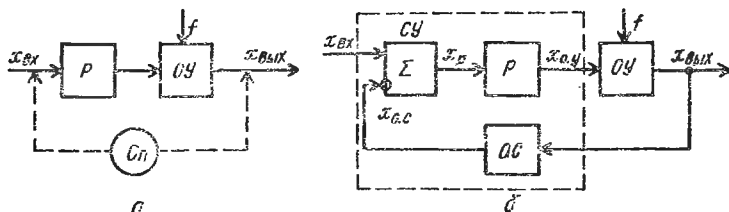


Рис. 1.2. Управление:

а — по разомкнутому циклу; б — по замкнутому

ваясь на выражении (1.2), то легко убедиться, что в этом случае управление невозможно: любое отклонение величины $x_{\text{вых}}$ от заданного будет усиливаться системой. Таким образом, главная обратная связь должна быть только отрицательной.

Процесс управления, осуществляемый без участия человека (рис. 1.2, б), называется *автоматическим*. Такая схема управления является типичной для автоматического управления любыми объектами, характер работы которых позволяет перевести их в автоматический режим.

Совокупность всех устройств, обеспечивающих автоматическое управление объектом, называется *системой автоматического управления* (САУ). В тех случаях, когда система обеспечивает стабилизацию управляемой величины в заданных пределах, она называется *системой автоматического регулирования* (САР). Очевидно, что автоматическое регулирование является частным видом автоматического управления. Последнее, однако, обеспечивает управление объектом по более сложным законам, чем стабилизация управляемой величины.

Из рис. 1.2, б видно, что САУ может быть определена как совокупность объекта управления и регулятора (обратная связь при этом рассматривается как часть регулятора). Сложность структуры регулятора в конкретной системе автоматического управления зависит от назначения объекта управления и требуемых режимов его работы. Однако в любом регуляторе должны содержаться источники информации о задачах и результатах управления, устройства, анализирующие информацию и вырабатывающие управляющие воздействия, исполнительные устройства, реализующие их. Схема системы управления, представленная как совокупность элементов, выполняющих определенное функциональное назначение, называется *функциональной схемой системы*. Представление систем управления в виде функциональных схем позволяет анализировать различные по физической сущности системы с позиций ряда общих требований.

Независимо от назначения системы, физической природы ее отдельных частей, их конструктивного исполнения функциональная схема системы управления в общем виде может быть представлена как совокупность следующих функциональных элементов (рис. 1.3): УЭ — управляющий (задающий) элемент; СУ — суммирующее устройство; КУ — последовательное корректирующее устройство; У — усилительный элемент; ИЭ — исполнительный элемент; МОС — местная обратная связь; ОУ — объект управления; ГОС — главная обратная связь.

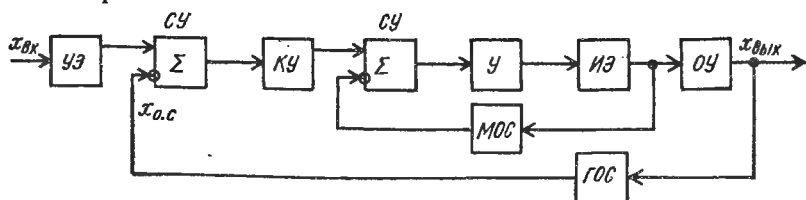


Рис. 1.3. Функциональная схема управления

Управляющее входное воздействие $x_{вх}$ прикладывается к управляющему элементу УЭ, выходная величина которого сравнивается с выходной величиной главной обратной связи $x_{о.с.}$. Результирующее воздействие корректируется в последовательном корректирующем устройстве КУ и после сравнения с величиной $x_{м.о.с.}$ — сигналом местной обратной связи — усиливается в усилительном устройстве У. Далее управляющее воздействие с помощью исполнительного элемента ИЭ передается на объект управления ОУ. Функции главной обратной связи ГОС были описаны выше.

Местная обратная связь МОС выполняет функции, аналогичные функциям ГОС, но эта связь охватывает не всю систему, а лишь часть ее, передавая изменения выходной величины некоторого промежуточного элемента системы на вход другого, предшествующего элемента. Если ГОС может быть только отрицательной, то МОС может быть как отрицательной, так и положительной; знак местной обратной связи и требуемый характер изменения величины $x_{м.о.с.}$ определяются в процессе анализа системы. Местную обратную связь называют также *параллельным корректирующим устройством* в тех случаях, когда она формируется по результатам динамического анализа системы, а не является функциональной частью самой системы.

Функциональные элементы конкретной системы управления конструктивно представляют собой различные механические, электрические, гидравлические, пневматические, электронные и другие устройства. При этом один и тот же функциональный элемент может включать в себя несколько конструктивных элементов и, наоборот, несколько функциональных элементов могут быть связаны в один конструктивный узел.

В качестве управляющих или задающих элементов используются, например, источники эталонного напряжения, эталонные сопротивления и т. п., воздействуя на которые можно устанавливать заданные значения переменных.

Главная обратная связь обычно включает в себя несколько конструктивных элементов, среди которых наиболее типовыми являются преобразователи-датчики, предназначенные для преобразования одной физической величины в другую, более удобную для контроля, измерения и дальнейшего использования в качестве управляющего сигнала. Преобразователями могут служить трансформаторы, выпрямители, инверторы, индуктивные элементы, датчики температуры, давления, частоты вращения (например, тахогенераторы) и др.

Последовательные и параллельные корректирующие устройства обеспечивают требуемые законы управления в системе. Эти устройства обычно вводятся в рабочую систему дополнительно по результатам анализа ее динамических свойств. Часто в качестве корректирующих устройств используются конструктивные элементы основной системы с изменением некоторых их параметров (сопротивлений, индуктивностей, емкостей и т. п.).

Усилительные устройства предназначены для усиления управляющих воздействий, которые чаще всего маломощны и не могут

непосредственно использоваться для целей управления. В качестве усилителей в системах используются электронные, электромашинные, магнитные, гидравлические и другие усилители.

Исполнительные элементы предназначены для приведения в движение регулирующих органов объекта управления, от положения которых зависят значения управляемых величин системы. В качестве исполнительных элементов широко используются, например, электродвигатели различных типов: двигатели постоянного тока, двухфазные и трехфазные асинхронные машины и др.

Помимо ручного и автоматического управления широко используется *автоматизированное управление*. На современном этапе развития теории управления под автоматизированным понимается управление объектом в системе с разомкнутой обратной связью с участием человека (людей), анализирующего состояние объекта с помощью электронно-вычислительных машин. Системы, реализующие такое управление, называются *автоматизированными системами управления* (АСУ). При этом, если объект управления относится к типу технических, то системы управления, построенные по указанному принципу, называются автоматизированными системами управления технологическими процессами. Если объект управления является объектом производственно-экономического или социального характера, то система управления им относится к автоматизированным системам организационного управления (АСОУ).

В последние годы все более широко внедряется интегрированное управление, реализуемое *интегрированными автоматизированными системами управления* (ИАСУ). В ИАСУ объектами управления являются технические, производственно-экономические, организационные и социальные системы. Так же, как и АСУ, ИАСУ создаются и функционируют на основе ЭВМ и экономико-математических методов, которые используются для управления техническими объектами, технологическими процессами, для планирования, контроля, анализа и регулирования производства в целом.

Общая схема процесса управления в АСУ и ИАСУ представлена на рис. 1.4. В этой схеме управляемому объекту (процессу, системе) задается определенное состояние $X_{\text{вых}}$, например требуемые параметры технологии, сроки и объемы выпуска продукции и т. п. Под влиянием внешней среды, в состав которой входят и вышестоящие органы управления, а также под влиянием процессов, проис-

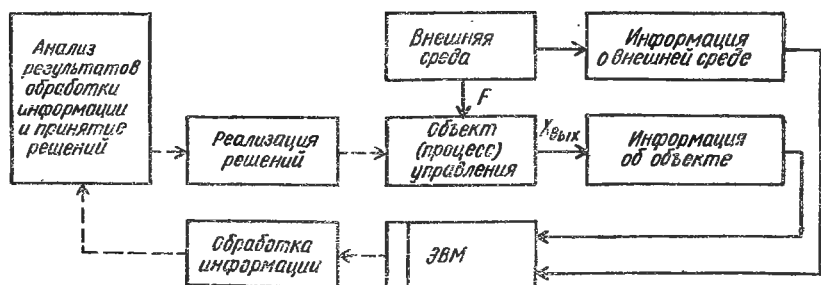


Рис. 1.4. Схема автоматизированного управления

ходящих внутри управляемой системы (износ оборудования, изменение квалификации персонала, морально-психологического климата и т. п.), объекты управления выходят из заданного состояния. Информация о состоянии внешней среды и о состоянии управляемой системы собирается, обрабатывается с помощью ЭВМ и анализируется с использованием экономико-математических моделей, размещенных в запоминающих устройствах ЭВМ. В результате анализа формируются управляющие воздействия, которые реализуются с помощью системы организационно-технических мероприятий.

При автоматизированном управлении техническими объектами и технологическими процессами сбор информации о состоянии объекта (процесса) и среды осуществляется с помощью датчиков и автоматической системы передачи информации от места ее возникновения к ЭВМ. При управлении производственно-экономическими и организационными системами сбор и передача информации осуществляется как с использованием датчиков, так и специально создаваемой системой сбора, регистрации и передачи производственно-экономической информации.

Сопоставляя рис. 1.4 и 1.2, можно заметить, что автоматизированное управление выполняется по разомкнутому циклу и с участием людей (операторов). Решение о выборе характера управляющего воздействия при автоматизированном управлении так же, как и при ручном, принимает человек (лицо, принимающее решение, — ЛПР). Однако, в отличие от ручного управления, автоматизированное управление формируется на базе ЭВМ и экономико-математических методов — аппарата, необходимого для анализа сложных систем.

Таким образом, автоматизированное управление можно определить как управление с разомкнутой обратной связью, в контуре которой используются ЭВМ и ЛПР.

Систему автоматизированного проектирования (САПР) можно определить как интегрированную автоматизированную систему управления, объектом управления которой является процесс выбора проектно-конструкторских решений на основе экономико-математических моделей изделий, конструкций, архитектурно-планировочных вариантов и т. п.

На схеме (рис. 1.5) представлены виды управления и типы систем, реализующих соответствующие виды управления. Следует от-

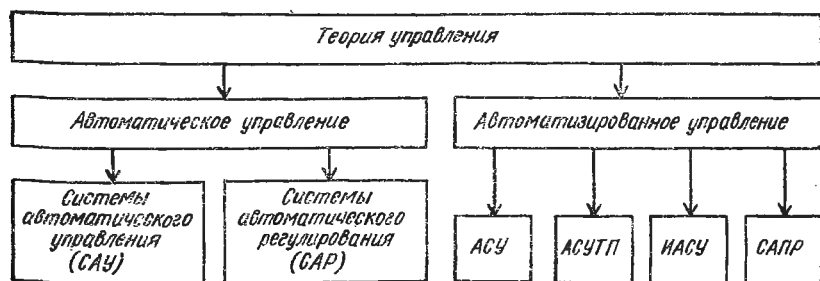


Рис. 1.5. Виды управления и типы систем управления

метить, что автоматическое и автоматизированное управление базируется на общих принципах и методах. Однако автоматизированное управление осуществляется в сложных (больших) системах, элементы которых значительно труднее описывать формализованно, чем элементы автоматических систем. Более того, объекты и процессы сложных систем зачастую вообще не могут быть описаны математически. Это потребовало более глубокой разработки методов и средств управления, реализуемого автоматизированными системами, а также создания ряда специфических методов анализа и синтеза сложных систем управления.

1.2. ПРИМЕРЫ ЭЛЕМЕНТОВ И СИСТЕМ УПРАВЛЕНИЯ

Выбирая в качестве примеров те или иные элементы и системы управления, будем исходить из того, насколько хорошо выбранные примеры позволяют иллюстрировать основные определения и понятия теории управления. С этих позиций наиболее удобными являются различные электромеханические устройства и системы, которые, не требуя специальных технических знаний, позволяют хорошо проследить процессы управления, его принципы и законы, режимы работы систем управления. В качестве первого простого примера объекта управления возьмем генератор постоянного тока независимого возбуждения (рис. 1.6,а), принцип действия которого хорошо известен из курса физики средней школы.

Пусть требуется управлять (регулировать) напряжением (ЭДС) генератора u , т. е. $x_{\text{вых}} = u$. Как известно, напряжение генератора зависит от частоты вращения его вала ω , т. е. от частоты вращения приводного двигателя (ПД) и от тока i_b в обмотке возбуждения ОВГ. Регулирование напряжения изменением частоты вращения ω , как правило, не применяется по технико-экономическим соображениям, но регулирование напряжения изменением тока возбуждения применяется широко, являясь в большинстве случаев единственным способом. Следовательно, $x_{\text{вх}} = i_b$.

В качестве второго примера объекта управления возьмем электродвигатель постоянного тока независимого возбуждения (рис. 1.6,б). Пусть требуется управлять частотой вращения ω вала двигателя. Не останавливаясь на всех возможных способах регулирования частоты вращения вала электродвигателя, отметим один

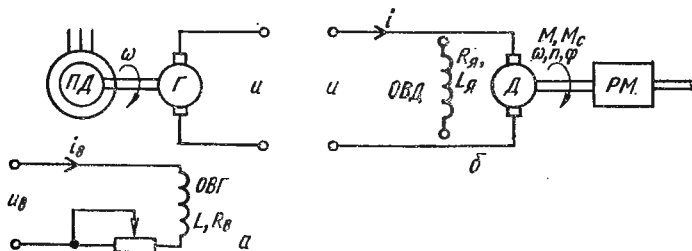


Рис. 1.6. Объекты управления:

а — электрический генератор постоянного тока; б — электрический двигатель постоянного тока независимого возбуждения

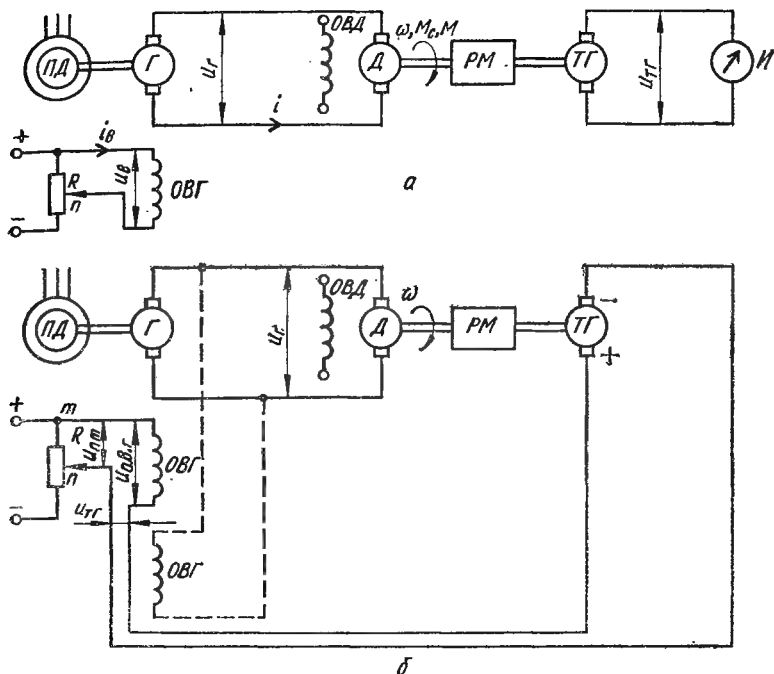


Рис. 1.7. Системы управления частотой вращения электродвигателя постоянного тока:
а — разомкнутая; б — замкнутая

из них — регулирование частоты вращения путем изменения напряжения, подводимого к якору, т. е. величины U . Следовательно, в рассматриваемом случае имеем $x_{ВХ} = U$; $x_{ВЫХ} = \omega$.

Как пример системы управления рассмотрим электромашинную схему управления частотой вращения двигателя постоянного тока (рис. 1.7). Предварительно рассмотрим процессы управления по разомкнутому циклу (рис. 1.7, а).

Объектом управления в схеме является электродвигатель постоянного тока Д с обмоткой независимого возбуждения двигателя ОВД. Электродвигатель приводит во вращение вал рабочей машины РМ, совершающей определенную рабочую операцию, например, передвижение строительного крана, рабочего органа землеройной машины и т. п. С валом электродвигателя жестко связан тахогенератор ТГ. Напряжение тахогенератора $U_{ТГ}$, пропорциональное частоте вращения электродвигателя, измеряется вольтметром, отградуированным в единицах частоты вращения ω . Электродвигатель Д подключен к генератору постоянного тока Г, вращаемому приводным двигателем ПД. Обмотка возбуждения генератора ОВГ через потенциометр R включена на источник постоянного напряжения. Перемещая ползунок n реостата и контролируя частоту вращения вала двигателя Д по показаниям измерителя И, можно управлять частотой вращения ω по заданному закону.

Очевидно, что для управления частотой вращения электродвигателя в рассматриваемой системе необходимо участие оператора, т. е. здесь имеет место ручное управление. Так как в рассмотренной системе изменение выходной величины ω непосредственно (без участия оператора) не преобразуется в управляющее воздействие на входе системы, то автоматическое управление в ней невозможно. При одном и том же положении регулирующего органа (ползунка n), но при разных значениях возмущающих воздействий на объект управления будут разными значения управляемой (регулируемой) величины — в данном случае частоты вращения ω вала двигателя Д.

Преобразуем разомкнутую систему управления (рис. 1.7, а) в замкнутую (рис. 1.7, б). С этой целью напряжение с выхода тахогенератора подадим на вход системы управления, обеспечив выполнение условия

$$u_{0.в.г} = u_{nm} - u_{тг}, \quad (1.3)$$

где $u_{0.в.г}$ — напряжение на обмотке возбуждения генератора; u_{nm} — напряжение, снимаемое с потенциометра.

В преобразованной таким образом системе изменения регулируемой величины — частоты вращения вала электродвигателя — автоматически передаются на вход и вместе с задающим напряжением u_{nm} формируют управляющее воздействие $u_{0.в.г}$. Выполнение условия (1.3) обеспечивает образование в системе главной отрицательной обратной связи и возможность автоматического регулирования выходной величины — частоты вращения ω .

Действительно, при увеличении нагрузки на валу электродвигателя частота вращения его уменьшается. Это вызывает уменьшение напряжения тахогенератора $u_{тг}$ и увеличение напряжения $u_{0.в.г}$ в соответствии с выражением (1.3). Возрастание величины $u_{0.в.г}$ приведет к возрастанию напряжения u_g генератора, подводимого к якору электродвигателя, что повлечет за собой восстановление значения частоты вращения вала электродвигателя.

Рассуждая аналогично, можно убедиться в том, что увеличение частоты вращения, вызванное уменьшением нагрузки на валу двигателя, также повлечет ее автоматическое восстановление до исходного значения. Следовательно, процесс стабилизации частоты вращения осуществляется автоматически, без участия оператора.

В рассматриваемой схеме обратная связь образована электрическим соединением тахогенератора с потенциометром и обмоткой возбуждения генератора. Используя определения, понятия и терминологию, введенные в предыдущем параграфе, по аналогии с рис. 1.2, б, в замкнутой системе (рис. 1.7, б) имеем:

- $x_{вх} = u_{nm} = u_z$ — задающее воздействие;
- $x_{о.с} = u_{тг}$ — сигнал обратной связи;
- $x_p = u_{0.в.г}$ — регулирующее воздействие, прикладываемое к регулятору;
- $x_{о.у} = u_g$ — управляющее воздействие, прикладываемое к объекту управления;

Основной составляющей возмущающего воздействия на замкнутую систему является нагрузка на валу рабочей машины, связанной с валом электродвигателя. Однако другие составляющие возмущения, выводящие систему из заданного состояния, могут быть приложены в любой точке системы, к любому ее элементу. Например, составляющими возмущения могут быть изменения: напряжения в сети, питающей приводной двигатель; сопротивления под щетками генератора и электродвигателя; сопротивления электрических цепей, входящих в систему, вызванные изменением температурного режима, и т. д. Система автоматического управления должна быть организована так, чтобы компенсировать влияние любого возмущения и обеспечить заданный режим работы объекта управления.

На рис. 1.7, б штриховой линией показана местная обратная связь, образованная с помощью второй обмотки возбуждения ОВГ, подключенной на выход генератора. Местная обратная связь, как отмечалось в предыдущем параграфе, может быть как положительной, так и отрицательной. В рассматриваемой системе при положительной местной обратной связи магнитные потоки обеих ОВГ складываются, при отрицательной — вычитаются, что обусловливается изменением полярности ОВГ, образующей обратную связь. Знак местной обратной связи определяется по условиям динамического расчета системы, о чем речь пойдет ниже.

По характеру действия своего выходного сигнала (выходной величины) местные обратные связи подразделяются на жесткие и гибкие. *Жесткая обратная связь* характерна тем, что сигнал на ее выходе существует как в установившемся режиме, т. е. в статике, так и в переходном режиме, т. е. в динамике. Действие жесткой обратной связи в системе проявляется независимо от того, какой сигнал имеется у нее на входе — постоянный или переменный. *Гибкая обратная связь* проявляется только в переходном режиме, т. е. в динамике. Очевидно, что обратная связь в схеме на рис. 1.7, б является жесткой. Гибкой ее можно сделать, подключив вторую обмотку возбуждения генератора на выходное напряжение генератора через емкость или трансформатор, что обеспечит передачу изменений входного сигнала местной обратной связи на выход этой связи лишь в переходных режимах.

Следует подчеркнуть, что главная обратная связь может быть только жесткой, в противном случае процесс управления выходной величиной в статике прекратится, даже если эта величина значительно превышает заданную.

Исходя из сказанного, схему на рис. 1.7, б можно определить как систему автоматического регулирования частоты вращения двигателя постоянного тока с главной обратной связью по частоте вращения и с местной жесткой обратной связью по напряжению генератора, питающего электродвигатель (или по напряжению, подводимому к двигателю). Все элементы, входящие в систему, кроме рабочей машины и электродвигателя Д, образуют регулятор.

Применяя к рассмотренной системе названия элементов, введенные при определении функциональной схемы САУ (рис. 1.3), отметим, что функции управляющего (задающего) элемента здесь выполняет потенциометр, а генератор объединяет в себе функции элемента сравнения, усилительного и исполнительного элементов; функции параллельного корректирующего устройства выполняет обмотка возбуждения генератора, подключенная на выходное напряжение генератора, а главная обратная связь образована цепью тахогенератора и ОБГ, включенной на потенциометр.

Система управления на рис. 1.7, б приведена в качестве простого и наглядного примера технической системы автоматического управления, на которой удобно проиллюстрировать ряд понятий теории управления, таких как объект управления, управляющие и управляемые величины, типы обратных связей, реализация функций управления с помощью конкретных физических устройств и т. д. Вместе с тем эта система широко применяется в различных отраслях техники, например для управления электроприводом строительных машин — экскаваторов, кранов и т. п. — во всех случаях, когда требуется широкое и плавное управление частотой вращения вала рабочего механизма. Система также находит широкое применение при управлении транспортными механизмами. Она позволяет при наличии нерегулируемого (или плохо регулируемого) и нереверсивного приводного двигателя ПД обеспечить широкое, плавное и реверсивное управление рабочей машины. Очевидные недостатки системы — низкий коэффициент полезного действия из-за неоднократного преобразования энергии в элементах системы, потребность в дополнительных машинах (приводной двигатель, генератор) — привели к тому, что область использования рассмотренной САУ значительно сузилась. Тем не менее эта система играет определенную роль в технике управления различными механизмами.

Технологию управления при автоматизированном управлении рассмотрим в общем виде, не конкретизируя объект управления. Как отмечалось, автоматизированное управление реализуется с помощью АСУ. Технология обработки информации и выработки управляющих воздействий в АСУ носит весьма типовой характер. Схема ее представлена на рис. 1.8.

Информацию об объекте управления в АСУ можно разделить на виды: условно-постоянную и оперативную.

Условно-постоянная информация содержится в документах, характеризующих состояние объекта в определенные периоды времени (производственные мощности предприятия, технические характеристики оборудования и т. п.). Эта информация необходима для составления машинной модели объекта, для чего документы преобразуются в форму, удобную для ввода в ЭВМ. В процессе подготовки для ввода в ЭВМ документы регистрируются, комплектуются и с использованием систем классификации и кодирования переносятся на машинные носители — перфокарты, перфоленты, магнитные ленты, магнитные карты. В зависимости от масштабов и назначения объекта, мощности вычислительного

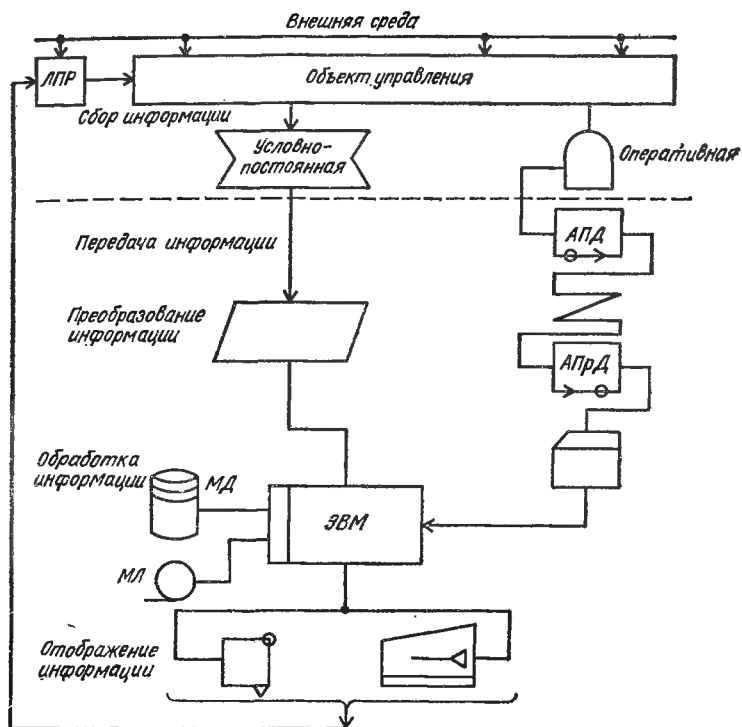


Рис. 1.8. Схема обработки информации и выработки управляющего решения в автоматизированной системе управления

центра и других факторов подготовка условно-постоянной информации для ввода в ЭВМ может осуществляться либо непосредственно на объекте, либо в информационно-вычислительном центре.

С машинных носителей информация вводится в ЭВМ и обрабатывается с использованием различных программ нормативно-справочных данных, математических моделей объектов и процессов, хранимых во внешних запоминающих устройствах ЭВМ — магнитных лентах МЛ, магнитных дисках МД и др.

Результаты обработки информации выводятся на устройства отображения информации — экранные пульты ЭП, алфавитно-цифровые печатающие устройства АЦПУ, пишущие машинки ПМ и др.

Оперативная информация характеризует показатели функционирования в текущий момент времени (состояние ресурсов, текущий объем выпуска продукции, потребление энергии и т. п.). Эта информация преобразуется в форму, удобную для передачи в ЭВМ по каналам связи, содержащим аппаратуру передачи данных АПД и аппаратуру приема данных АПРД. Результаты обработки оперативной информации также выводятся на средства отображения.

Анализируя результаты обработки информации, ЛПР — лицо, принимающее решение, формирует управляющие воздействия на объект управления, которыми могут быть те или иные административные распоряжения, поставки дополнительных ресурсов, указания и т. п.

Рассмотренная схема обработки информации и выработки управляющих решений является типичной для большинства АСУ, независимо от характера объекта. При этом в автоматизированных системах управления технологическими процессами (АСУТП) для сбора оперативной информации используются датчики, контролирующие состояние оборудования.

Из приведенного описания видно, что для любой АСУ характерны следующие этапы обработки информации об объекте управления: сбор, контроль и регистрация, перевод информации на машинные носители, ввод информации в ЭВМ, обработка на ЭВМ, вывод результатов обработки информации на устройства отображения и хранения.

1.3. ПРИНЦИПЫ И ЗАКОНЫ УПРАВЛЕНИЯ (РЕГУЛИРОВАНИЯ)

Технико-экономические показатели систем управления существенно зависят от принципов и законов управления (регулирования), реализуемых той или иной системой.

Принцип управления можно рассматривать как способ формирования управляющего воздействия, а закон управления — как математическое выражение принципа. Остановимся подробнее на этих понятиях, так как они являются фундаментальными в теории управления.

Системы управления формируются на основе одного из трех принципов: 1) управление по отклонению управляемой величины; 2) управление по возмущению; 3) комбинированное управление.

В системах управления по отклонению отклонение управляемой величины от заданного значения преобразуется в управляющее воздействие. Это преобразование осуществляется регулятором, передается на объект управления и направлено на уменьшение отклонения управляемой величины от заданного значения. Такой принцип управления известен также под названием принципа Ползунова — Уатта.

Способ реализации принципа регулирования по отклонению рассмотрим на примере центробежного регулятора частоты вращения вала рабочей машины РМ (рис. 1.9), широко применяемого в системах управления современными механизмами. Клапан 1, регулирующий подачу топлива из канала А к двигателю (канал В), связан через тягу 2, рычаг CD и систему шарнирных соединений с грузиками 3.

Валик 5 связан с валом рабочей машины РМ, вращение которого через шарнирную передачу 4 передается системе грузиков 3. Пружина 6 играет роль чувствительного элемента регулятора. При заданной частоте вращения ω вала рабочей машины система от-

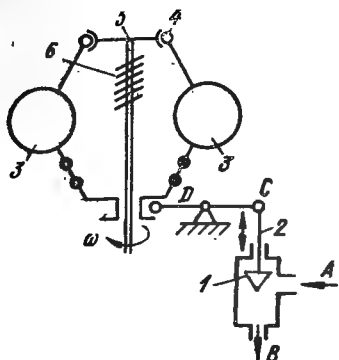


Рис. 1.9. Регулирование по отклонению

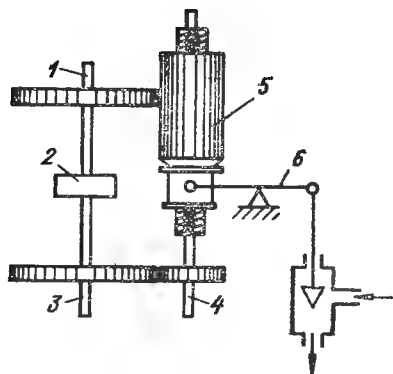


Рис. 1.10. Регулирование по возмущению

регулирована так, чтобы положение клапана 1 обеспечивало требуемую подачу топлива к двигателю. Если частота вращения превысит заданную, то под влиянием центробежной силы грузики 3 расходятся, конец рычага D поднимается, а конец C — опускается. Клапан 1 займет новое положение, при котором подача топлива к двигателю уменьшится и частота вращения вала РМ примет вновь заданное значение. Рассуждая аналогично, можно убедиться в том, что при уменьшении частоты вращения ω , по сравнению с заданной, клапан переместится в направлении, обеспечивающем увеличение подачи топлива к двигателю и восстановление частоты вращения.

Нетрудно убедиться в том, что система регулирования скорости вращения электродвигателя, рассмотренная в предыдущем параграфе (см. рис. 1.7, б), является системой регулирования по принципу отклонения регулируемой величины от заданного значения. В этой системе значение регулируемой величины — частоты вращения ω вала рабочей машины РМ — задается положением ползунка n потенциометра R . При отклонении ω от заданного значения изменяется напряжение тахогенератора $u_{тг}$, и в системе формируется управляющее воздействие, направленное на приведение регулируемой величины ω к заданному значению.

Достоинством систем регулирования по отклонению является то, что в них обеспечивается учет всех возмущений, действующих на любой элемент системы, так как управляющее воздействие формируется в зависимости от конечного результата управления — значения регулируемой величины.

Недостатком регулирования по отклонению является снижение быстродействия системы, выражающееся в замедлении реакции регулятора на изменение регулируемой величины. Действительно, при регулировании по отклонению регулируемой величины компенсация влияния любого возмущения, независимо от места его приложения в системе, начинается лишь после того, как это воздействие преобразовалось элементами системы в изменение регулируемой величины.

Отмеченный недостаток устраняется в системах управления по принципу компенсации возмущения — управления по возмущению.

При управлении по возмущению в системе измеряется возмущающее воздействие (нагрузка) и результат измерения преобразуется в управляющее воздействие. Реализацию принципа регулирования по возмущению рассмотрим на примере кинематической схемы регулятора, представленного на рис. 1.10. В этой схеме вал двигателя 1 и вал рабочей машины 3 связаны между собой упругим элементом 2. Вращение валов 1 и 3 передается валу 4, верхняя часть которого имеет червячную передачу. Муфта 5 наряду с вращательным движением может осуществлять возвратно-поступательное движение вдоль линии вала 4, которое с помощью рычага 6 передается в устройство регулирования подачи топлива, описанного выше (см. рис. 1.9). Из рис. 1.10 видно, что при изменении нагрузки на валу РМ произойдет относительное смещение сечений валов 1 и 3, смещение муфты 5 вдоль линии вала 4 и соответствующее изменение подачи энергии к двигателю: при снижении нагрузки подача энергии уменьшится, а при повышении — увеличится.

Сложность измерения возмущения (нагрузки) — один из недостатков систем регулирования по возмущению. Однако в ряде технических устройств часто бывает нетрудно выделить и использовать с целью регулирования некоторую физическую величину, связанную определенным законом (например, пропорциональную) с возмущением. Примером таких устройств служит генератор постоянного тока со смешанным возбуждением, у которого последовательная (токовая) обмотка может быть использована для формирования управляющего воздействия, зависящего от нагрузки — тока якоря генератора.

Достоинством систем регулирования по возмущению является их более высокое быстродействие, так как система реагирует на изменение возмущения до того, как эти изменения элементами системы преобразуются в изменения регулируемой величины.

Существенным недостатком регулирования по возмущению является нечувствительность систем ко всем возмущениям, кроме того, на которое настроен регулятор. В связи с этим в практике управления широко используется принцип комбинированного управления, когда для формирования управляющего воздействия используются как изменения возмущения, так и изменения регулируемой величины. Примером системы комбинированного управления может служить рассмотренная выше схема (рис. 1.7, б), если с помощью второй обмотки возбуждения генератора осуществить связь по току в цепи двигателя, так как ток двигателя связан с нагрузкой на валу рабочей машины.

Описанная выше (см. рис. 1.4) автоматизированная система управления является системой комбинированного управления, так как в ней используется принцип регулирования по отклонению регулируемых величин на основе анализа информации об объекте управления и принцип регулирования по возмущению на основе анализа информации о состоянии среды.

Каждый из трех описанных принципов управления может быть реализован по различным законам. Законом управления назовем функциональную связь между воздействием Y регулятора на регулируемый орган и значениями величин, определяющих принцип управления.

Обозначим отклонение управляемой величины от заданного значения через $\epsilon = \Delta x$, а нагрузку — через q . Тогда в наиболее общем виде законы управления выразятся так:

$$Y = F(\epsilon, \epsilon', \epsilon'', \dots, \int \epsilon dt, \dots, q, q', \dots, \int q dt). \quad (1.4)$$

Выбор того или иного закона управления производится на основе анализа динамических режимов системы управления, а реализуется выбранный закон с помощью технических средств, вводимых в систему в качестве корректирующих устройств. Примером таких устройств являются пассивные RC - и RL -контуры, стабилизирующие трансформаторы и др. Действительно, если на первичную обмотку стабилизирующего трансформатора подать электрическое напряжение u_1 , пропорциональное корректируемому сигналу управления, то при определенном подборе параметров первичной и вторичной цепей на выходе трансформатора (на выходе вторичной обмотки) возникает напряжение u_2 , пропорциональное первой производной первичного напряжения: $u_2 \approx \frac{du_1}{dt}$.

Таким образом, если $u_1 = \epsilon$; $u_2 = y$, то получим $y = \epsilon'$, реализуя один из законов управления, содержащийся в выражении (1.4), а именно: управление по первой производной от отклонения управляемой величины.

Укажем некоторые простейшие законы управления, вытекающие из выражения (1.4):

закон управления, пропорциональный отклонению (управление по отклонению управляемой величины от заданного значения): $y = a_0 \epsilon$;

пропорциональный отклонению с воздействием по первой производной от отклонения: $y = a_0 \epsilon + a_1 \epsilon'$;

пропорциональный отклонению с воздействием по первой и второй производным от отклонения: $y = a_0 \epsilon + a_1 \epsilon' + a_2 \epsilon''$;

интегральный по отклонению: $y = b_1 \int_0^t \epsilon dt$;

пропорциональный отклонению управляемой величины с воздействием интеграла по нагрузке: $y = a_0 \epsilon + b_1 \int_0^t q dt$;

пропорциональный нагрузке: $y = b_0 q$.

Количество законов управления из бесконечного множества законов, определяемых выражением (1.4), ограничено возможностями технических и организационно-технических средств, реализующих тот или иной закон. Реальное техническое средство обычно не позволяет реализовать выбранный закон управления (например, ввести первую производную от корректируемой величины) в чистом виде без генерирования помех. При этом количество по-

мех и сложность их устранения возрастают с усложнением законов управления.

Остановимся подробнее на двух законах управления, которые наиболее часто встречаются в технических системах наряду с пропорциональными законами управления по отклонению управляемой величины от заданного значения и по возмущению (нагрузке). Это законы управления с введением производных и интегралов по отклонению управляемой величины и по нагрузке.

При управлении по отклонению к объекту управления подводится управляющее воздействие, пропорциональное отклонению, т. е.

$$y = a_0 \varepsilon, \quad (1.5)$$

где a_0 — коэффициент пропорциональности между отклонением и управляющим воздействием.

Вводя в закон (1.5) производную по отклонению управляемой величины, получим

$$y = a(\varepsilon + a_1 \varepsilon'), \quad (1.6)$$

где a , a_1 — соответствующие коэффициенты пропорциональности.

Управление по закону (1.6) придает системе управления иные динамические свойства, чем управление по закону (1.5). Действительно, пусть отклонение управляемой величины от заданного значения происходит по кривой 1 (рис. 1.11, а). Задача системы управления заключается в том, чтобы возможно быстрее компенсировать отклонение или свести его к допустимому значению, определенному требованиями к процессу управления.

При управлении по закону (1.5) величина ε передается на вход системы через главную обратную связь с отрицательным знаком, осуществляя управляющее воздействие, направленное на уменьшение отклонения. Управляющее воздействие в этом случае пропорционально ординате кривой 1 в каждый момент времени. Поскольку обратная связь отрицательна, то в промежутке $0 < t < t_1$, т. е. в течение времени, когда $\varepsilon > 0$, управляющее воздействие направлено на уменьшение отклонения. В силу инерционности элементов системы характер управляющего воздействия сохранится какое-то время при $t > t_1$, хотя знак ε уже изменился. Это вызовет нарас-

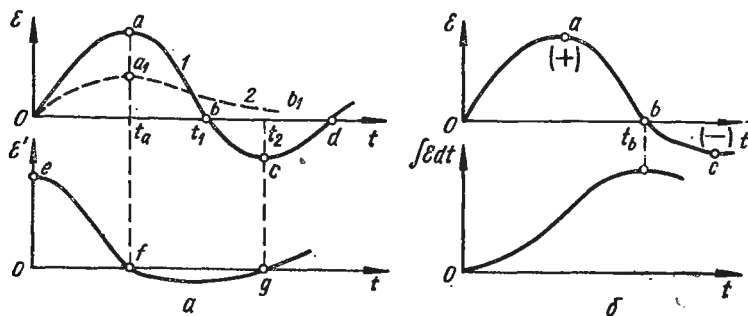


Рис. 1.11. Законы управления:
а — по производной; б — по интегралу

тание отклонения в обратную сторону, которое прекратится в некоторый момент времени $t_2 > t_1$. На участке cd при $t > t_2$ регулятор будет действовать в сторону уменьшения ϵ и т. д. Как видно, регулятор запаздывает в своей реакции на изменение отклонения, что приводит к колебаниям управляемой величины и к увеличению времени ее стабилизации.

Показатели процесса управления существенно улучшаются при переходе к регулированию по закону (1.6). Действительно, в этом случае на участке $0a$ кривой I воздействие регулятора, направленное на уменьшение отклонения, возрастает, так как к управляющей составляющей, пропорциональной ϵ , добавляется составляющая, пропорциональная ϵ' . При этом на участке $0a$ обе составляющие имеют одинаковый знак, и регулятор значительно эффективнее, чем раньше, сдерживает нарастание отклонения. На участке abc кривой I знак составляющей управления, пропорциональной ϵ , сохранился, а знак составляющей управления, пропорциональной ϵ' , изменился на обратный, так как в точке a изменился знак производной функции $\epsilon(t)$.

Таким образом, составляющая, пропорциональная производной от отклонения, форсирует действие регулятора, направленное на уменьшение отклонения до тех пор, пока оно возрастает, и поддерживает значение отклонения в период его уменьшения. Это сокращает время регулирования, предотвращает колебания выходной величины.

Составляющая, пропорциональная производной, проявляется тем эффективнее, чем быстрее нарастает отклонение. При этом еще до того, как отклонение накопится, создается эффективное воздействие для его уменьшения, так как управляющий эффект, обеспечиваемый этой составляющей, зависит не от величины отклонения, а от скорости его нарастания.

Следует подчеркнуть, что регулирование только по производной без составляющей, пропорциональной отклонению, невозможно, так как в этом случае процесс управления будет проявляться только в динамике, прерываясь в статических режимах. Коррекция только по производной может осуществляться посредством местной обратной связи, а главная обратная связь должна содержать составляющую, пропорциональную отклонению.

При введении составляющей управления, пропорциональной интегралу от отклонения, имеем:

$$y = a(\epsilon + b \int_0^t \epsilon dt). \quad (1.7)$$

Вторая составляющая, пропорциональная интегралу, нарастает все время, пока отклонение положительно (рис. 1.11, б). Следовательно, на участке $0 < t < t_b$ регулирующее воздействие регулятора увеличивается даже при незначительном значении отклонения в каждый момент времени. Это повышает точность систем управления. В то же время регулирование по интегралу способствует развитию в системе колебательных процессов. Действительно, при $t = t_b$, когда отклонение изменяет свой знак, составляющая, про-

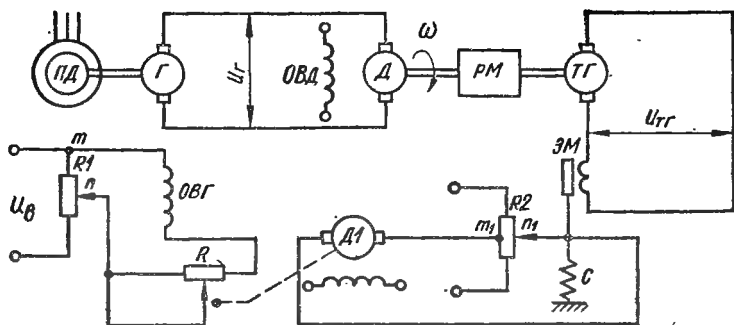


Рис. 1.12. Астатическое регулирование

порциональная интегралу, выражаясь суммарной (с учетом знака) площадью под кривой $\varepsilon(t)$, начинает уменьшаться, но по знаку не изменяется. Знак этой составляющей изменится лишь тогда, когда площадь, ограниченная нижней ветвью кривой $\varepsilon(t)$, превысит площадь, ограниченную верхней ее ветвью. До этого момента, начиная с момента $t=t_b$, регулятор будет способствовать увеличению отрицательного отклонения. Продолжая рассуждения, можно заметить, что составляющая управления, пропорциональная интегралу от отклонения, придает системе колебательность.

Очевидно, что при использовании закона регулирования по интегралу можно создавать системы управления, в которых отклонения регулируемых величин от заданных сводятся к нулю. Такое управление, т. е. управление без отклонения управляемой величины от заданного значения (или управление без ошибки), называется *астатическим*. Управление же, при котором в установившихся режимах работы системы сохраняется отклонение управляемой величины от заданного значения, называется *статическим*.

В системе автоматического регулирования, представленной на рис. 1.7, б, процесс регулирования статический, а сама система — статическая система автоматического регулирования. Отклонение в такой системе неизбежно, так как оно обеспечивает управляющий эффект в ней. Чтобы превратить эту систему в астатическую, ее необходимо дополнить устройствами, вводящими в закон управления составляющую, пропорциональную интегралу от отклонения.

Астатическая система регулирования частоты вращения двигателя постоянного тока приведена на рис. 1.12. Обозначение и назначение элементов в этой схеме такое же, как и на рис. 1.7. Кроме того, в обратную связь по частоте вращения включены электромагнит ЭМ, якорь которого удерживается пружиной С, потенциометр R1 и электродвигатель Д1, вал которого связан с ползунок реостата R в цепи ОВГ.

При частоте вращения ω вала рабочей машины РМ, равной заданной, система настраивается так, что напряжение, снимаемое с потенциометра R2 и подводимое к двигателю Д1, равно нулю: ползунок n_1 находится против средней точки m_1 потенциометра R2.

При возрастании или уменьшении частоты вращения ω равновесие между усилием пружины C и усилием, удерживающем якорь электромагнита, нарушается. Ползунок n_1 поднимается вверх (при возрастании ω , т. е. при возрастании $u_{тг}$) или опускается вниз (при уменьшении ω , т. е. при уменьшении $u_{тг}$). Таким образом, к двигателю Д1 будет подано напряжение того или иного знака. В результате ползунок реостата R в цепи ОБГ переместится в таком направлении, чтобы изменить напряжение u_r настолько, насколько требуется для восстановления заданного значения ω . Как видно из рисунка, процесс регулирования будет продолжаться до тех пор, пока отклонение ω от заданного значения полностью не устранится, так как только в этом случае прекратится вращение вспомогательного двигателя Д1 и перемещение ползунка реостата R .

Из приведенного примера видно, что придание системе астатических свойств существенно усложнило ее, сделало более дорогой и менее надежной. Кроме того, как отмечено выше, в системе могут возникнуть нежелательные колебания выходной величины. Поэтому астатическое регулирование осуществляется в тех случаях, когда необходимо полностью устранить отклонение регулируемой величины от требуемого значения, даже за счет усложнения системы и ухудшения некоторых динамических свойств системы.

1.4. РЕЖИМЫ РАБОТЫ СИСТЕМ УПРАВЛЕНИЯ

Для наиболее полного анализа работы систем управления целесообразно рассматривать их в двух режимах — статическом и динамическом.

Статический режим работы системы характеризуется тем, что величины, определяющие поведение системы, не изменяются во времени и система находится в равновесном состоянии. Раздел теории управления, в котором изучается равновесное состояние, называется *статикой*. Уместно подчеркнуть, что понятие «статика системы» и «статическое регулирование» — различные понятия. Первое определяет равновесное состояние системы, второе — процесс регулирования, при котором неизбежно отклонение регулируемой величины от заданной в равновесном состоянии. Из сказанного ясно, что понятие «статика» относится как к статическим, так и к астатическим системам управления.

Динамический режим характеризуется тем, что величины, определяющие поведение системы, изменяются во времени, т. е. динамика рассматривает переходные процессы.

Введем понятие «звено» системы управления, под которым будем понимать физический элемент системы, его часть или совокупность элементов, характеризующихся с точки зрения их свойств в статических и в динамических режимах. Одно из важнейших свойств звеньев — направленность их действия. Это свойство заключается в том, что изменение величины на входе звена влечет за собой изменение величины на его выходе, но не наоборот. Характер функциональной зависимости выходной величины от вход-

ной определяется статическими и динамическими свойствами звеньев.

Зависимость выходной величины звена от входной в равновесном состоянии называется *статической характеристикой звена*. В практике анализа систем управления широко используются как аналитические, так и графические представления статических характеристик в виде функций $x_{\text{вых}} = f(x_{\text{вх}})$, построенных в равновесных состояниях системы.

Статические характеристики реальных звеньев всегда нелинейны, более того, в ряде случаев их невозможно описать аналитически. Примером такой характеристики является зависимость индукции от напряженности — кривая намагничивания $B = f(H)$. Однако при упрощенном анализе многие звенья систем управления могут быть описаны линейными зависимостями вида

$$x_{\text{вых}} = a + kx_{\text{вх}},$$

где a и k — постоянные величины.

В таких случаях широко используются уравнения звеньев в приращениях входных и выходных величин:

$$\Delta x_{\text{вых}} = k \Delta x_{\text{вх}}.$$

Отношение приращения выходной величины звена к приращению входной в равновесном состоянии (в статике) называется *коэффициентом усиления*, или *коэффициентом передачи звена*:

$$k = \Delta x_{\text{вых}} / \Delta x_{\text{вх}}. \quad (1.8)$$

Коэффициент передачи звена определяется конструктивными параметрами и играет важную роль в характеристике звена, определяя его статические и динамические свойства.

Если статическая характеристика звена нелинейна, то коэффициент передачи k выражается как тангенс угла наклона касательной к кривой $x_{\text{вых}} = f(x_{\text{вх}})$ и будет различным для разных значений входной и выходной величин (рис. 1.13).

При рассмотрении астатического регулирования было показано, что при подаче на вход астатического звена приращения входной величины выходная величина этого звена неограниченно растет. Действительно, если на рис. 1.12 в равновесном состоянии системы отключить обмотку электромагнита ЭМ от тахогенератора и дать приращение напряжения на входе этой обмотки, то угол поворота вала двигателя Д1 будет непрерывно возрастать. Из ска-

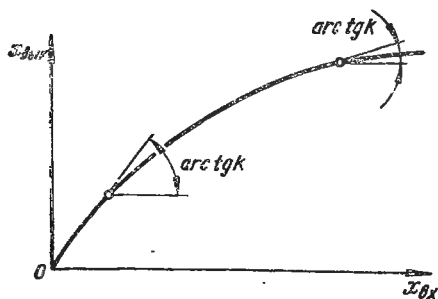


Рис. 1.13. Статическая характеристика нелинейного звена

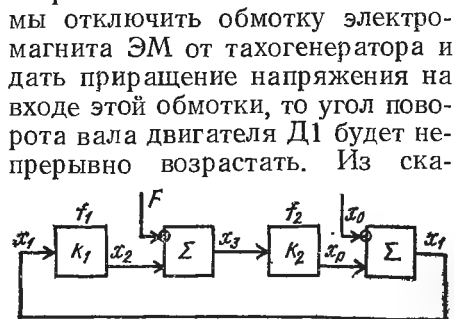


Рис. 1.14. Замкнутая одноконтурная система

званного видно, что выражение (1.8) не годится для определения понятия коэффициента передачи астатического звена. *Коэффициентом передачи (усиления) астатического звена* называется отношение скорости изменения выходной величины к приращению входной величины в установившемся состоянии:

$$k_a = \frac{\Delta x_{\text{вых}}/\Delta t}{\Delta x_{\text{вх}}}.$$

В ряде случаев установившегося состояния не имеет не только выходная величина астатического звена, но и ее первая, вторая и т. д. производные. В таких случаях статические характеристики астатических звеньев строятся как зависимости от входной величины второй, третьей и т. д. производных выходной величины. Сами астатические звенья в этих случаях называются *звеньями с астатизмом второго, третьего и т. д. порядка*.

Статические характеристики отдельных звеньев либо задаются как характеристики физических элементов, либо могут быть легко построены по паспортным данным элементов системы. Однако для анализа свойств системы управления необходимо располагать не только статическими характеристиками отдельных звеньев, но и статической характеристикой всей системы. В теории автоматического управления разработаны различные методы и приемы построения статических характеристик САУ на основе статических характеристик отдельных звеньев, образующих систему. При этом под статической характеристикой разомкнутой системы понимается зависимость выходной величины от входной в равновесном состоянии, а под статической характеристикой замкнутой системы понимается зависимость регулируемой величины от возмущающего воздействия в равновесном состоянии.

В большинстве случаев анализа САУ представляют интерес режимы работы замкнутой, а не разомкнутой системы. В связи с этим рассмотрим общий метод построения статической характеристики замкнутой системы. Для реализации метода необходимо замкнутую систему представить как одноконтурную (рис. 1.14), состоящую из двух частей: f_1 , объединяющую все звенья системы, находящиеся между регулируемой величиной x_p и возмущающим воздействием F , и f_2 , объединяющую все звенья системы, находящиеся между возмущающим воздействием и регулируемой величиной. При этом принимается во внимание отмеченное ранее свойство направленности звеньев САУ. Величина x_0 на рис. 1.14 означает заданное значение регулируемой величины.

Зависимости $x_2 = f_1(x_1)$ и $x_p = f_2(x_3)$ строятся на основании статических характеристик звеньев, образующих части системы f_1 и f_2 соответственно. Задача сводится к построению зависимости $x_p = f(F)$ — статической характеристики замкнутой системы.

Запишем уравнения статики рассматриваемой системы:

$$x_p = f_2(x_3); \quad x_3 = x_2 - F; \quad x_2 = f_1(x_1); \quad x_1 = x_0 - x_p.$$

Исключая из этих уравнений промежуточные величины, можно найти зависимость $x_p = f(F)$. При этом в случае линейных систем

нетрудно получить аналитическую зависимость статической характеристики $x_p = f(F)$. Действительно, для линейной системы имеем:

$$x_p = k_2 x_3; \quad x_3 = x_2 - F; \quad x_2 = k_1 x_1; \quad x_1 = x_0 - x_p,$$

где k_1 и k_2 — коэффициенты передачи частей системы f_1 и f_2 соответственно. Исключая из написанных уравнений промежуточные переменные, получим уравнение статической характеристики рассматриваемой системы:

$$x_p = \frac{k}{k+1} x_0 - \frac{k_2}{k+1} F = f(F), \quad (1.9)$$

где $k = k_1 k_2$ — коэффициент передачи разомкнутой системы.

Анализ выражения (1.9) показывает, что точность регулирования повышается с увеличением коэффициента передачи разомкнутой системы. Действительно, чем выше значение k , тем меньше значение x_p отличается от x_0 . Следует, однако, подчеркнуть, что увеличение k с целью повышения точности системы должно происходить за счет повышения коэффициента k_1 , а не k_2 , так как только в этом случае в выражении (1.9) уменьшается влияние второго слагаемого, определяющего влияние возмущающего воздействия на значение регулируемой величины. Наконец, чрезмерное увеличение коэффициента передачи разомкнутой системы может привести к нарушению устойчивости системы в динамических режимах работы, что будет рассмотрено ниже.

Определение точности регулирования является основной задачей анализа статических режимов САУ. Точность САУ определяется статической ошибкой регулирования Δ , под которой понимается разность между заданным значением регулируемой величины x_0 и ее действительным значением x в установившемся состоянии: $\Delta = x_0 - x$. Отношение статической ошибки Δ к номинальному значению регулируемой величины $x_{ном}$ называется *статизмом регулирования*: $\delta = \Delta / x_{ном}$.

Из рассмотрения процесса управления в замкнутой системе видно, что статическая ошибка в ней будет меньше, чем в разомкнутой системе. Действительно, уменьшение выходной величины в замкнутой системе повлечет за собой увеличение управляющего воздействия, направленного на восстановление выходной величины, т. е. на уменьшение статической ошибки. Не останавливаясь на более подробном обосновании, приведем соотношение между ошибкой регулирования в разомкнутой системе Δ_0 и ошибкой в замкнутой системе Δ :

$$\Delta = \Delta_0 / (k + 1), \quad (1.10)$$

где k — коэффициент усиления разомкнутой системы.

Как видно из изложенного, анализ статических режимов САУ базируется на анализе алгебраических уравнений статики и сводится, в конечном счете, к определению статической ошибки. Не являясь сам по себе сложным, статический расчет САУ играет очень важную роль. Если параметры системы, определяемые в результате статического расчета, будут выбраны неверно, то после-

дующие динамические расчеты САУ, будучи более трудоемкими, чем статические, окажутся напрасными.

Динамические режимы САУ анализируются на основе дифференциальных уравнений, описывающих переходные процессы в элементах и системах. Динамические свойства САУ оцениваются по количественным показателям переходных процессов в системе, обусловленных некоторыми типовыми возмущающими воздействиями, прикладываемыми к системе. Характер типового воздействия выбирается из следующих соображений. Во-первых, это воздействие должно в максимальной степени отражать реальные возмущения, возможные в проектируемой системе. Во-вторых, это воздействие должно быть сравнительно легко воспроизводимым как при теоретическом, так и при экспериментальном исследовании. В-третьих, количественные показатели переходных процессов, обусловленных типовыми воздействиями, должны легко пересчитываться в показатели переходных процессов, обусловленных возмущениями любого характера. Таким требованиям в значительной степени отвечает типовое воздействие в виде мгновенного скачка входной величины — единичного скачка. Действительно, единичное воздействие на входе в большой мере отражает процессы пуска и торможения в реальной системе и легко реализуется как аналитически, так и экспериментально.

Показатели динамики САУ определяются по форме зависимости $x_{\text{вых}}(t)$ при подаче на вход единичного скачка $x_{\text{вх}} = [1]$ (рис. 1.15). К числу основных показателей динамических режимов САУ относятся устойчивость, время переходного процесса, перерегулирование, колебательность, среднеквадратичная ошибка и др.

Устойчивостью называется способность системы управления, выведенной из равновесного состояния, с течением времени возвратиться вновь в равновесное состояние. Кривая 1 на рис. 1.15 характеризует устойчивую систему, а кривая 2 — неустойчивую. Устойчивость — основное требование, предъявляемое к САУ, так как неустойчивая система неработоспособна и должна быть стабилизирована. Стабилизация САУ достигается соответствующим изменением ее параметров без ущерба для основного назначения системы, призванной выполнять определенные функции. Подробнее вопросы анализа САУ на устойчивость освещены ниже.

Временем переходного процесса t_n в САУ называется отсчитываемое от начала приложения воздействия время, в течение которого регулируемая величина достигает значения, отличающегося от установившегося в заданных пределах. На рис. 1.15 знаком Δ обозначена допустимая статическая ошибка регулирования.

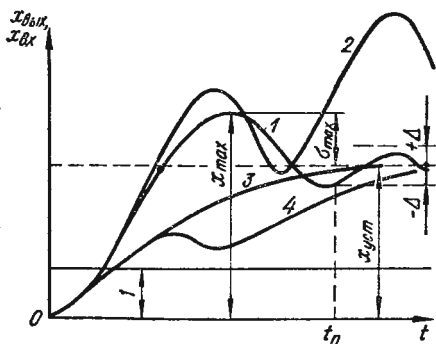


Рис. 1.15. Основные показатели динамики САУ

Перерегулированием $\sigma_{\max}$ называется максимальное отклонение регулируемой величины от установившегося значения. Обычно эта величина выражается в процентах от установившегося значения регулируемой величины:

$$\sigma_{\max} = \frac{x_{\max} - x_{уст}}{x_{уст}} \cdot 100 \, \%.$$

Колебательность определяется числом μ переходов регулируемой величины через установившееся значение (числом колебаний) в течение времени t_n . При $\mu > 2$ переходный процесс определяется как колебательный, при $\mu = 0$ — как процесс без перерегулирования (кривые 3 и 4). Если при $\mu = 0$ величина $\dot{x}_{\text{вых}} > 0$, то переходный процесс характеризуется как монотонный (кривая 3 на рис. 1.15).

Среднеквадратичная ошибка определяется как положительный квадратный корень: $\Delta_{\text{ср.кв}} = M[|x^*(t) - x(t)|^2]$, где $x(t)$, $x^*(t)$ — желаемое и фактическое значения выходной величины; M — символ математического ожидания.

Среднеквадратичная ошибка позволяет оценить динамическую точность системы.

Перечисленные показатели динамических режимов САУ часто называют показателями качества системы. Значения этих показателей содержатся в задании на проектирование системы управления.

1.5. ПОСТАНОВКА И МЕТОДЫ РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ

Постановка задач управления, в общем виде не зависит от характера объекта управления, но методы решения этих задач существенно отличаются.

Во всех случаях задача управления заключается в том, чтобы в любой момент времени обеспечить требуемое состояние объекта управления. Это состояние определяется набором функциональных характеристик.

Различают задачи управления: 1) в технических системах; 2) в организационно-экономических системах.

Технические системы часто называют динамическими, имея в виду высокую скорость протекания переходных процессов в них. Задача управления объектом в технической системе может быть поставлена либо как задача анализа системы управления, либо как задача синтеза.

При решении *задачи анализа* известна структура всей системы и параметры объекта управления, определяемые из назначения и условий эксплуатации объекта. Задача разработки системы управления в этом случае сводится к определению параметров регулятора, обеспечивающих заданные статические и динамические показатели системы управления.

Исходными данными для решения задачи анализа систем управления являются их математические модели, составляемые на

основе математического описания элементов и системы в целом. Математическое описание динамических систем управления базируется на алгебраических и дифференциальных уравнениях, а также на разностных уравнениях. Для решения дифференциальных уравнений динамики САУ широко используется преобразование Лапласа, позволяющее перейти от функций времени $x(t)$ к функциям комплексного переменного $x(p)$:

$$x(p) = \int_0^{\infty} x(t) e^{-pt} dt, \quad (1.11)$$

где p — комплексное число.

Функция $x(t)$ в выражении (1.11) называется *оригиналом*, а $x(p)$ — *изображением*.

В задачах анализа линейных систем автоматического управления преобразование Лапласа осуществляется при нулевых начальных условиях, т. е. при условиях, когда при $t=0$ сама функция $x(t)$ и ее производные обращаются в нуль. В таких случаях замена оригиналов изображениями позволяет перейти от дифференциальных уравнений динамики САУ к алгебраическим уравнениям. Действительно, пусть имеется функция

$$\varphi(t) = \frac{d^n f(t)}{dt^n}. \quad (1.12)$$

Применив к выражению (1.12) преобразование (1.11) при нулевых начальных условиях, найдем

$$\varphi(p) = p^n f(p). \quad (1.13)$$

Таким образом, при указанных условиях переход от дифференциальных уравнений оригиналов к зависимостям в форме изображений сводится к тому, что в исходных дифференциальных уравнениях динамики САУ оригиналы функций заменяются их изображениями, а операции дифференцирования — умножением на p . В итоге система дифференциальных уравнений заменяется системой алгебраических уравнений, которые решаются значительно проще. В результате решения получаем изображения, от которых можно вновь перейти к оригиналам путем обратного преобразования с использованием теорем операционного исчисления. Часто, однако, обратного преобразования не требуется, так как в теории управления разработаны эффективные методы определения качества управления, базирующиеся на анализе операторных уравнений — уравнений в изображениях. Часть этих методов рассматривается в дальнейшем.

Хотя описанные постановки и метод решения задач анализа САУ охватывают линейные системы, на базе этих методов в значительной степени базируются исследования других типов систем управления: нелинейных, дискретных, производственно-экономических.

Задача синтеза систем управления заключается в определении ее структуры и параметров по заданной цели управления. Обычно синтез системы управления сводится к синтезу регулятора, так

как структура и параметры объекта управления известны или определяются специалистами соответствующих областей на основе общего назначения проектируемой системы.

Повышение требований к технико-экономическим показателям систем управления привело к созданию оптимальных систем, под которыми понимаются системы, обеспечивающие наилучшее в некотором смысле управление. Синтез систем управления также проводится на основании их математических моделей.

Производственно-экономические и организационные системы являются сложными системами. Постановка и решение задач управления в таких системах отличается большей сложностью, вытекающей из их математического описания. Объекты и процессы в этих системах, как правило, не могут быть достаточно полно описаны математически. Многие процессы, протекающие в сложных системах, могут быть оценены не количественно, а качественно, для чего используется аппарат булевых переменных.

Булевы переменные представляют собой качественные параметры, оцениваемые по двухбалльной системе. Они удобны для качественной оценки событий типа ДА и НЕТ, при этом наступление события оценивается баллом 1, а ненаступление — баллом 0.

Функциональные зависимости, описывающие элементы сложных систем, содержат как непрерывные, так и булевы переменные, например:

$$x = f(y, \beta),$$

где x и y — непрерывные переменные; β — булев параметр.

Аналитические зависимости $x = f(y)$ задаются отдельно для значений $\beta = 0$ и $\beta = 1$.

Так же, как и в динамических системах, в сложных системах соотношения между количественными параметрами формализуются в виде алгебраических, дифференциальных, разностных уравнений, а соотношения между качественными параметрами устанавливаются в виде логических соотношений, в форме таблиц и т. д. При описании сложных систем широко используются вероятностные функции, системы весов, оценивающие те или иные события, экспертные оценки.

Сложность математического описания производственно-экономических и организационных систем вытекает прежде всего из того, что это человеко-машинные системы, функционирующие часто в условиях трудно предсказуемых ситуаций. Процессы в таких системах весьма сложно описать формализованно.

Кроме того, поведение сложной системы обуславливается огромным количеством переменных, исчисляемых сотнями тысяч. Обычно эти величины имеют разную физическую природу, соотношения между ними отличаются исключительным разнообразием. Переменные, характеризующие сложную систему, находятся в тесном взаимодействии и взаимосвязи, характер которых таков, что изменение отдельной связи или параметров какого-либо элемента, входящего в систему, влечет за собой изменение всех других свя-

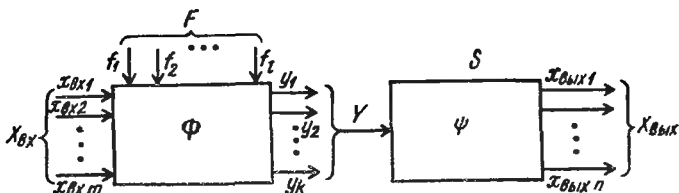


Рис. 1.16. Управление в сложной системе

зей и параметров или большинства из них. Всякий новый элемент, вводимый в сложную систему, изменяет параметры и связи многих или всех элементов системы.

Исходя из сказанного, сформулируем проблему управления в сложной системе в общем виде и укажем на одну из существенных особенностей решения этой проблемы.

Пусть имеет сложную систему S (рис. 1.16), состоящую из двух частей — Φ и Ψ . Часть Φ сложной системы преобразует множество управляющих (входных) воздействий $X_{вх} = \{x_{вх1}, x_{вх2}, \dots, x_{вхn}\}$ и множество возмущающих воздействий $F = \{f_1, f_2, \dots, f_i\}$ в переменные, характеризующие состояние системы (переменные состояния): $Y = \{y_1, y_2, \dots, y_k\}$. Часть Ψ преобразует переменные состояния в выходные величины $X_{вых} = \{x_{вых1}, x_{вых2}, \dots, x_{выхn}\}$.

Сложность проблемы управления заключается в том, что каждая переменная множества Y — множества координат состояния системы — является функцией всех воздействий на систему:

$$y_i = \Phi_i(X_{вх}, F), \quad i = \overline{1, k},$$

а каждая переменная множества $X_{вых}$ — множества управляемых (выходных) переменных — является функцией всех переменных состояния:

$$x_{выхi} = \Psi_i(Y), \quad i = \overline{1, n}$$

или

$$x_{выхi} = S_i(X_{вх}, F), \quad i = \overline{1, n}.$$

Задача управления сложной системой сводится к тому, чтобы обеспечить такие функциональные преобразования Φ , Ψ , S , которые были бы оптимальными по выбранным критериям эффективности — некоторым показателям, определяющим степень достижения цели, стоящей перед управляемой системой.

Для упрощения математического описания сложных систем их разбивают на подсистемы по принципу иерархии, устанавливают связи между элементами каждого уровня и между отдельными уровнями.

Характерной особенностью проблемы управления сложными системами является необходимость системного подхода к управлению. Он заключается в том, что система должна рассматриваться как единое целое с позиций цели функционирования, общей для всех подсистем. Практически это приводит к тому, что недопус-

тима независимая оптимизация функционирования отдельных подсистем, образующих систему, т. е. недопустимо усовершенствование отдельных подсистем сложной системы с позиций частных целей этих подсистем. Например, в системе управления каким-либо производственным объектом недопустимо совершенствование управления отдельным подразделением без учета целей, стоящих перед объектом в целом. В противном случае, например, цех какого-либо предприятия существенно улучшит свои показатели, а работа предприятия в целом может оказаться неудовлетворительной.

Для иллюстрации сказанного рассмотрим некоторую условную систему, находящуюся под влиянием двух управляющих воздействий — y_1 и y_2 , которые являются аргументами функций φ и f . Пусть функция φ отражает доходы в системе, а f — потери: $\varphi = F_1(y_1, y_2)$; $f = F_2(y_1, y_2)$. В качестве критерия эффективности такой системы целесообразно выбрать следующий:

$$\max (\varphi - f) = \max [F_1(y_1, y_2) - F_2(y_1, y_2)].$$

Допустим, что одна подсистема максимизирует доходы, т. е. максимизирует φ по y_1 , а другая минимизирует убытки, т. е. минимизирует f по y_2 :

$$\varphi_1 = \max_{y_1} [F_1(y_1, y_2)]; \quad f_1 = \min_{y_2} [F_2(y_1, y_2)].$$

Очевидно, что $\max (\varphi - f) \neq \varphi_1 - f_1$, т. е.

$$\max (\varphi - f) \neq \max_{y_1} [F_1(y_1, y_2)] - \min_{y_2} [F_2(y_1, y_2)].$$

Таким образом, управление сложной системой должно осуществляться как управление единой целостной структурой, имеющей конечную цель функционирования, общую для всех подсистем, образующих систему.

Структура сложной системы управления определяется в результате системного анализа, представляющего собой совокупность приемов исследования сложных систем, объединяющую как формализованные, так и неформализованные методы. Формализованные методы базируются на прикладных математических дисциплинах, прежде всего на методах математического программирования. Неформализованные методы опираются на экспертные оценки и аппарат теории полезности, позволяющий упорядочивать критерии управления по их весу в оценке степени достижения цели управления. В качестве технической базы системного анализа используются ЭВМ.

При известной структуре сложной системы цель ее функционирования описывается некоторой скалярной функцией — целевой функцией W , достигающей экстремального значения при оптимальном (наилучшем в определенном смысле) управлении. Аргументами целевой функции являются выходные параметры $y_1, y_2, \dots, y_n$, которые в реальной системе отражают эффективность ее функционирования. Выходные параметры — аргументы целевой функции — сами являются функциями некоторых величин $x_1, x_2, \dots, x_m$, кото-

рые в реальной системе являются управляемыми параметрами. Воздействуя на них, можно обеспечить требуемые значения выходных параметров.

Условия управления в сложной системе зависят также от управляемых параметров $\alpha_1, \alpha_2, \dots, \alpha_l$.

В процессе функционирования сложной системы на управляемые, неуправляемые и выходные параметры накладываются ограничения, отражающие реальные условия работы и существования системы. Ограничениями в реальных системах могут быть, например, ограничения тех или иных ресурсов, сроков производства работ, стоимости, геометрических размеров и т. д.

На основании сказанного математическая модель сложной системы в общем виде может быть представлена так:

$$W = W(y_1, y_2, \dots, y_n); \quad (1.14)$$

$$y_i = y_i(x_1, x_2, \dots, x_m; \alpha_1, \alpha_2, \dots, \alpha_l); \quad (1.15)$$

$$f_j(x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n; \alpha_1, \alpha_2, \dots, \alpha_l) \{ \cong \} 0. \quad (1.16)$$

Функция W , выражающая цель управления, называется *целевой функцией*. Так как ее аргументы являются функциями многих переменных, то функцию W называют еще *функционалом*. Зависимость (1.16) выражает ограничения, накладываемые на параметры, обусловленные конкретными условиями функционирования сложной системы.

В изложенной постановке задача управления сложной системой заключается в том, чтобы обеспечить экстремум целевой функции: $W \rightarrow \text{extr}$. При этом в зависимости от смысла целевой функции задача управления сводится к обеспечению $W \rightarrow \min$ или $W \rightarrow \max$. Очевидно, что в первом случае целевая функция отражает затраты в системе (например, минимум себестоимости выпускаемой продукции), а во втором — прибыль, увеличение выпуска продукции и т. д. Решение поставленной задачи обеспечивается соответствующим подбором значений управляемых параметров. Огромное число управляемых и выходных параметров, сложные зависимости (1.14) — (1.16), включающие существенные нелинейности, дискретность некоторых параметров и другие факторы делают решение задачи управления весьма сложной.

Основными методами решения задач управления в изложенной постановке являются методы математического программирования. Математическое программирование представляет собой раздел прикладной математики, занимающийся вопросами определения экстремумов функций на некотором множестве.

При решении задач управления сложными системами методы математического программирования используются для определения численных значений переменных, характеризующих производственно-экономические процессы, удовлетворяющих некоторым ограничениям в форме уравнений или неравенств и обеспечивающих экстремум заданной целевой функции, которая отражает эффективность управления. В зависимости от вида целевой функции и ограничений используются различные виды математического про-

граммирования — линейное, динамическое, выпуклое, геометрическое и др. Особенности этих методов и область их применения рассмотрены ниже.

Задачи управления сложными системами существенно усложняются влиянием многообразных случайных факторов, например выходом из строя оборудования, погодными условиями, нарушением поставок сырья или комплектующих изделий и т. п. В таких случаях управление должно осуществляться в условиях неполной информации об объекте управления. Эти задачи решаются методами стохастического программирования. Широко используются различные эвристические методы.

Глава 2

ЛИНЕЙНЫЕ ДИНАМИЧЕСКИЕ СИСТЕМЫ УПРАВЛЕНИЯ

2.1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ЗВЕНЬЕВ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

Линейными называются системы управления, которые и в статике и в динамике описываются линейными уравнениями. Одна из основных особенностей линейных систем заключается в том, что к ним применим принцип суперпозиции, в соответствии с которым реакция системы на совокупность возмущений определяется суммой реакций на каждое возмущение, прикладываемое к системе в рассматриваемый момент времени.

Реальные системы управления, особенно сложные, — существенно нелинейны. К линейному описанию можно условно свести только небольшой класс динамических систем, но и в этих случаях свойства реальных систем лишь приближенно отображаются их линейными моделями. Однако, несмотря на ограниченность линейных моделей, их роль в теории управления очень велика. Это объясняется тем, что допущение о линейности систем управления в ряде случаев не приводит к недопустимым погрешностям, с одной стороны, а с другой — существенно упрощает исследование систем. Кроме того, методы исследования реальных нелинейных систем управления в значительной степени базируются на методах исследования линейных систем.

Теория управления линейными системами разработана достаточно глубоко и располагает эффективными и простыми методами анализа и синтеза систем управления, в основном систем автоматического управления и регулирования. Следует подчеркнуть, что линейная теория управления позволяет изучать линейные модели реальных процессов и объектов, а не сами процессы и объекты. Динамические режимы линейных систем исследуются с помощью их математических моделей. При этом любую динамическую линейную систему можно представить в виде совокупности следу-

ющих типовых структурных звеньев: аperiodических, колебательных, интегрирующих, дифференцирующих и усилительных. Каждое из названных звеньев достаточно полно характеризуется формой дифференциального уравнения, видом передаточной и видом переходной функции.

Дифференциальное уравнение звена определяет связь между его выходными и входными величинами в динамических режимах.

Передаточная функция звена $W(p)$ представляет собой отношение преобразованной по Лапласу выходной величины звена к преобразованной по Лапласу входной величине при нулевых начальных условиях:

$$W(p) = x_{\text{вых}}(p)/x_{\text{вх}}(p). \quad (2.1)$$

Переходная функция $h(t)$ показывает характер переходного процесса в звене, на входе которого приложено единичное входное воздействие:

$$h(t) = x_{\text{вых}}(t) \text{ при } x_{\text{вх}}(t) = [1].$$

Указанные характеристики звена составляют важные элементы аппарата исследования динамических линейных систем. Рассмотрим эти характеристики для каждого из перечисленных звеньев.

Аperiodическое звено

Аperiodическим называется звено, в котором связь между выходной и входной величинами выражается уравнением

$$T \frac{dx_{\text{вых}}}{dt} + x_{\text{вых}} = kx_{\text{вх}}, \quad (2.2)$$

где k — коэффициент усиления (передачи) звена; T — постоянная времени звена, с. Величины k и T выражаются через физические параметры конкретного звена.

Применяя к уравнению (2.2) преобразование (1.11) — преобразование Лапласа при нулевых начальных условиях, получим передаточную функцию аperiodического звена:

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{k}{Tp + 1}. \quad (2.3)$$

Решая уравнение (2.2) относительно $x_{\text{вых}}(t)$, получим

$$x_{\text{вых}}(t) = kx_{\text{вх}}(1 - e^{-t/T}) \quad (2.4)$$

или при $x_{\text{вх}} = 1$ имеем переходную функцию аperiodического звена:

$$h(t) = k(1 - e^{-t/T}). \quad (2.5)$$

График переходного процесса в аperiodическом звене представлен на рис. 2.1, а кривой 1. Как видно из уравнения (2.4) и рис. 2.1, а, переходный процесс в аperiodическом звене полностью определяется значениями коэффициента усиления звена k и его постоянной времени T .

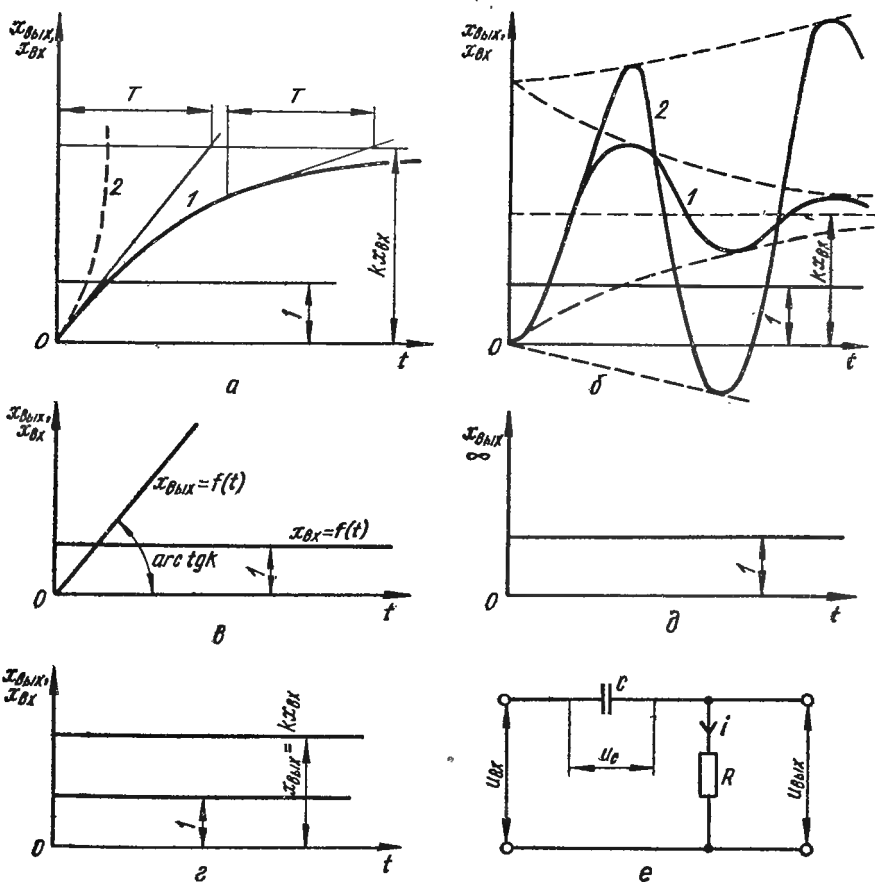


Рис. 2.1. Переходные процессы в звеньях:

а — аperiodическом; б — колебательно; в — интегрирующем; г — усилительном д — идеальном дифференцирующем; е — схема дифференцирующего звена

Если дифференциальное уравнение аperiodического звена имеет вид

$$T \frac{dx_{\text{вых}}}{dt} - x_{\text{вых}} = kx_{\text{вх}}, \quad (2.6)$$

то переходный процесс в нем описывается уравнением

$$x_{\text{вых}}(t) = ke^{t/T} \quad (2.7)$$

и представляется кривой 2 на рис. 2.1, а. Звено, описываемое уравнением (2.6), называется *неустойчивым* аperiodическим звеном.

Аperiodические звенья в линейных динамических системах встречаются очень часто. Приведем некоторые примеры.

Пример 1. Пусть к обмотке ОВГ (рис. 1.6, а) подан скачок напряжения $u_{\text{в}}$. Дифференциальное уравнение рассматриваемого звена имеет вид

$$u_{\text{в}} = i_{\text{в}} R_{\text{в}} + L \frac{di_{\text{в}}}{dt} \quad \text{или} \quad \frac{u_{\text{в}}}{R_{\text{в}}} = i_{\text{в}} + T \frac{di_{\text{в}}}{dt},$$

где $T = L/R_v$ — постоянная времени цепи ОВГ; L — индуктивность цепи; R_v — сопротивление цепи.

Учитывая, что в рассматриваемом звене $i_v = x_{\text{вых}}$, $u_v = x_{\text{вх}}$, получаем

$$T \frac{dx_{\text{вых}}}{dt} + x_{\text{вых}} = kx_{\text{вх}},$$

где $k = x_{\text{вых}}/x_{\text{вх}} = 1/R_v$ — коэффициент усиления.

Пример 2. Если в схеме (рис. 1.6, б) рассмотреть связь между переменными ω и i , то, используя электромеханические свойства рассматриваемой системы при условии пропорциональности между ω и M_c — моментом сопротивления на валу электродвигателя, можно получить

$$T \frac{d\omega}{dt} + \omega = ki,$$

где $T = I/c_1$ — постоянная времени; I — момент инерции, приведенный к валу двигателя; c_1 — коэффициент пропорциональности между M_c и ω ; $k = k_m/c_1$ — коэффициент усиления; k_m — коэффициент пропорциональности между моментом M , развиваемым двигателем, и током i .

Колебательное звено

Колебательным называется звено, в котором связь между выходной и входной величинами выражается уравнением

$$T^2 \frac{d^2 x_{\text{вых}}}{dt^2} + 2\xi T \frac{dx_{\text{вых}}}{dt} + x_{\text{вых}} = kx_{\text{вх}} \quad (2.8)$$

при условии $\xi < 1$. В уравнении (2.8) T — постоянная времени; k — коэффициент усиления; ξ — коэффициент затухания.

Решение линейного дифференциального уравнения (2.8), а следовательно, характер изменения $x_{\text{вых}}(t)$ зависит от значения корней соответствующего характеристического уравнения

$$T^2 \alpha^2 + 2\xi T \alpha + 1 = 0; \quad (2.9)$$

$$\alpha_{1,2} = -\frac{1}{T} (\xi \pm \sqrt{\xi^2 - 1}). \quad (2.10)$$

При $\xi < 1$ корни α_1 и α_2 — комплексные. В этом случае переходный процесс в звене носит колебательный характер, а переходная функция колебательного звена имеет вид

$$h(t) = k \left[1 - \frac{e^{-\xi t/T}}{\sqrt{1-\xi^2}} \sin \left(\frac{\sqrt{1-\xi^2}}{T} t + \arctg \frac{\sqrt{1-\xi^2}}{\xi} \right) \right]. \quad (2.11)$$

Колебания (2.11) носят затухающий характер — кривая 1 на рис. 2.1, б. Действительно, из (2.11) при $t \rightarrow \infty$ имеем $x_{\text{вых}}(t) \rightarrow k$.

Применяя к уравнению (2.8) преобразование Лапласа при нулевых начальных условиях, получим передаточную функцию устойчивого колебательного звена:

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{k}{T^2 p^2 + 2\xi T p + 1}. \quad (2.12)$$

Если дифференциальное уравнение звена имеет вид

$$T^2 \frac{d^2 x_{\text{ВЫХ}}}{dt^2} - 2\xi \frac{dx_{\text{ВЫХ}}}{dt} + x_{\text{ВЫХ}} = kx_{\text{ВХ}}, \quad (2.13)$$

то переходная функция

$$h(t) = k \frac{e^{\xi t/T}}{\sqrt{1-\xi^2}} \sin \left(\frac{\sqrt{1-\xi^2}}{T} t + \arctg \frac{\sqrt{1-\xi^2}}{\xi} \right), \quad (2.14)$$

Из (2.14) при $t \rightarrow \infty$ следует $h(t) \rightarrow \infty$, т. е. колебания в таком звене носят расходящийся характер (кривая 2 на рис. 2.1, б). Звено, в котором связь между входной и выходной величинами описывается дифференциальным уравнением (2.13) при $\xi < 1$, называется *неустойчивым колебательным звеном*.

Наконец, если в уравнении (2.8) $\xi > 1$, то корни характеристического уравнения (2.9) будут вещественными:

$$\alpha_1 = -\frac{\xi + \sqrt{\xi^2 - 1}}{T}; \quad \alpha_2 = -\frac{\xi - \sqrt{\xi^2 - 1}}{T}.$$

В этом случае

$$h(t) = k \left(1 - \frac{T_1}{T_1 - T_2} e^{-t/T_1} + \frac{T_2}{T_1 - T_2} e^{-t/T_2} \right), \quad (2.15)$$

где $T_1 = -1/\alpha_1$, $T_2 = -1/\alpha_2$.

Таким образом, при $\xi > 1$ уравнение (2.8) описывает два апериодических звена, соединенных последовательно, имеющих постоянные времена T_1 и T_2 и коэффициенты усиления, произведение которых равно k .

Примером колебательного звена может являться двигатель постоянного тока независимого возбуждения (рис. 1.6, б), у которого $x_{\text{ВХ}} = u$ — напряжение, подводимое к якору электродвигателя; $x_{\text{ВЫХ}} = n$ — скорость вращения выходного вала, а момент сопротивления на валу $M_c = 0$, т. е. двигатель работает вхолостую, при этом учитывается индуктивность цепи якоря. При указанных условиях уравнение двигателя

$$T_M T_J \frac{d^2 n}{dt^2} + T_M \frac{dn}{dt} + n = ku, \quad (2.16)$$

где T_M — электромеханическая постоянная времени, характеризующая механическую инерцию вала; T_J — электромагнитная постоянная времени цепи якоря двигателя, характеризующая электромагнитную инерцию цепи якоря; k — коэффициент усиления.

Величины T_M , T_J и k определяются через параметры двигателя, в частности $T_J = L_J/R_J$, где L_J , R_J — индуктивность и сопротивление цепи якоря.

Обозначая в (2.16) $T_M T_J = T^2$; $T_M = 2\xi T$, получаем типовое уравнение колебательного звена, в котором $x_{\text{ВХ}} = u$; $x_{\text{ВЫХ}} = n$:

$$T^2 \frac{d^2 n}{dt^2} + 2\xi T \frac{dn}{dt} + n = ku.$$

Уравнением типа (2.8) описывается движение массы, подвешенной на пружине, электромагнитные процессы в электрической цепи, содержащей индуктивность L , активное сопротивление R , емкость C и многие другие звенья динамических систем. Все звенья такого типа имеют передаточную функцию вида (2.12). При этом величины k и T выражаются через конструктивные параметры соответствующего звена.

Интегрирующее звено

Интегрирующим называется звено, в котором выходная величина пропорциональна интегралу от входной величины:

$$x_{\text{вых}} = k \int_0^t x_{\text{вх}} dt \quad (2.17)$$

или

$$\frac{dx_{\text{вых}}}{dt} = k x_{\text{вх}}. \quad (2.18)$$

Применяя к (2.17) преобразование Лапласа при нулевых начальных условиях, получим передаточную функцию интегрирующего звена:

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{k}{p}. \quad (2.19)$$

На основании (2.17) имеем переходную функцию интегрирующего звена (рис. 2.1, в)

$$h(t) = kt.$$

Из (2.18) видно, что в интегрирующем звене скорость изменения выходной величины пропорциональна входной величине, т. е. интегрирующее звено является *астатическим*.

Примером интегрирующего звена может явиться двигатель (рис. 1.6, б) в котором в качестве входной величины рассматривается скорость вращения вала n , а в качестве выходной — угол его поворота φ . В этом случае имеем:

$$n = c \frac{d\varphi}{dt} \quad \text{или} \quad \varphi = \frac{1}{c} \int_0^t n dt = k \int_0^t n dt,$$

где k и c — коэффициенты пропорциональности.

Усилительное звено

Усилительным называется звено, в котором выходная величина пропорциональна входной:

$$x_{\text{вых}} = k x_{\text{вх}}; \quad W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = k. \quad (2.20)$$

Усилительное звено безынерционно — переходный процесс в нем отсутствует: выходная величина изменяется вместе с измене-

ниями входной величины, без сдвига во времени (рис. 2.1, *з*). В действительности любое реальное звено обладает инерционностью. Поэтому в динамической системе усилительным (безынерционным) принимается такое звено, в котором переходные процессы протекают несоизмеримо быстрее, чем в других звеньях системы. Примером усилительных звеньев может служить электронный усилитель в системах регулирования механических, тепловых и других инерционных процессов.

Дифференцирующее звено

Идеальным дифференцирующим называется звено, в котором выходная величина пропорциональна производной от входной величины:

$$x_{\text{ВЫХ}} = \tau \frac{dx_{\text{ВХ}}}{dt}, \quad (2.21)$$

где τ — постоянная времени звена, определяемая через его параметры.

Передаточная функция идеального дифференцирующего звена

$$W(p) = \frac{x_{\text{ВЫХ}}(p)}{x_{\text{ВХ}}(p)} = \tau p. \quad (2.22)$$

Из (2.22) видно, что при скачкообразном изменении входной величины значение выходной величины стремится к бесконечности, т. е. при $x_{\text{ВХ}} = [1]$; $x_{\text{ВЫХ}} = \infty$ (рис. 2.1, *д*). Естественно, что в реальных звеньях такой переходный процесс невозможен и его описание в форме (2.21) идеализировано.

В качестве дифференцирующего звена широко применяется *RC-контур* (рис. 2.1, *е*), для которого на основании законов Ома и Кирхгофа можно записать:

$$u_{\text{ВХ}} = u_c + iR = u_c + u_{\text{ВЫХ}} = \frac{1}{C} \int i dt + u_{\text{ВЫХ}}.$$

Учитывая, что $i = u_{\text{ВЫХ}}/R$, имеем:

$$u_{\text{ВХ}} = \frac{1}{\tau} \int u_{\text{ВЫХ}} dt + u_{\text{ВЫХ}},$$

где $\tau = RC$ — постоянная времени электрического контура, с.

Из последнего выражения получим

$$\tau \frac{du_{\text{ВЫХ}}}{dt} + u_{\text{ВЫХ}} = \tau \frac{du_{\text{ВХ}}}{dt}. \quad (2.23)$$

Принимая во внимание, что $u_{\text{ВЫХ}} = x_{\text{ВЫХ}}$; $u_{\text{ВХ}} = x_{\text{ВХ}}$, на основании (2.23) запишем

$$\tau \frac{dx_{\text{ВЫХ}}}{dt} + x_{\text{ВЫХ}} = \tau \frac{dx_{\text{ВХ}}}{dt}. \quad (2.24)$$

Подбирая параметры звена так, чтобы $\tau \frac{dx_{\text{ВЫХ}}}{dt} \ll x_{\text{ВЫХ}}$, из (2.24)

получим уравнение идеального дифференцирующего звена — уравнение (2.21).

На выходе реальных дифференцирующих звеньев помимо составляющей, пропорциональной производной от входной величины, генерируются также другие составляющие. В линейных динамических системах наряду с идеальным дифференцирующим звеном, описываемым уравнением (2.21) и передаточной функцией (2.22), в качестве типовых структурных звеньев приняты также:

реальное дифференцирующее звено первого порядка, описываемое уравнением

$$k \left(\tau \frac{dx_{\text{вх}}}{dt} + x_{\text{вх}} \right) = x_{\text{вых}}, \quad (2.25)$$

дифференцирующее звено второго порядка, описываемое уравнением

$$k \left(\tau^2 \frac{d^2 x_{\text{вх}}}{dt^2} + 2\xi\tau \frac{dx_{\text{вх}}}{dt} + x_{\text{вх}} \right) = x_{\text{вых}}. \quad (2.26)$$

Применяя к (2.25) и (2.26) преобразование Лапласа при нулевых начальных условиях, получим передаточную функцию дифференцирующего звена первого порядка

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = k(\tau p + 1) \quad (2.27)$$

и передаточную функцию дифференцирующего звена второго порядка

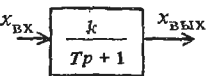
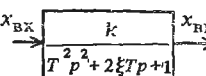
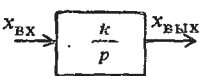
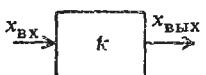
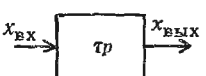
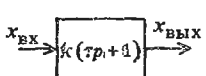
$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = k(\tau^2 p^2 + 2\xi\tau p + 1). \quad (2.28)$$

Постоянная времени τ и коэффициент усиления звена k определяются на основании параметров конкретных звеньев.

Дифференцирующие звенья широко используются в качестве корректирующих устройств, вводимых в систему для улучшения ее динамических свойств. С помощью таких звеньев в законы управления вводятся составляющие, пропорциональные производным по времени от отклонения или возмущения, что увеличивает быстродействие системы.

Описанными типовыми структурными звеньями охватываются все звенья, возможные в динамической системе управления. Нетрудно заметить универсальность приведенного математического описания. Действительно, описание звеньев динамических систем с использованием аппарата передаточных функций, базирующегося на исходных дифференциальных уравнениях, не зависит от их физической природы. Любая линейная система управления, независимо от назначения, структуры, физической природы ее элементов может быть представлена математической моделью в виде совокупности рассмотренных выше типовых структурных звеньев (табл. 2.1). Это является наглядным свидетельством единства материального мира: «Единство природы обнаруживается в поразительной аналогичности дифференциальных уравнений, относящихся

Таблица 2.1.

Звено	Передаточная функция	Типовое звено	Переходный процесс
Апериодическое	$W(p) = \frac{k}{Tp + 1}$		$x_{\text{ВЫХ}} = kx_{\text{ВХ}}(1 - e^{-t/T})$
Колебательное	$W(p) = \frac{k}{T^2 p^2 + 2\xi Tp + 1}$		$x_{\text{ВЫХ}} = kx_{\text{ВХ}} \left[1 - \frac{e^{-\xi/T t}}{\sqrt{1 - \xi^2}} \times \right. \\ \left. \times \sin\left(\frac{\sqrt{1 - \xi^2}}{T} t + \arctg \frac{\sqrt{1 - \xi^2}}{\xi}\right) \right]$
Интегрирующее	$W(p) = \frac{k}{p}$		$x_{\text{ВЫХ}} = kx_{\text{ВХ}} t$
Усилительное	$W(p) = k$		$x_{\text{ВЫХ}} = kx_{\text{ВХ}}$
Дифференцирующие	$W(p) = \tau p$		$x_{\text{ВЫХ}} = \tau \frac{dx_{\text{ВХ}}}{dt}$
	$W(p) = k(\tau p + 1)$		$x_{\text{ВЫХ}} = k\left(\tau \frac{dx_{\text{ВХ}}}{dt} + x_{\text{ВХ}}\right)$

ся к разным областям явлений» (Ленин В. И. Материализм и эмпириокритицизм // Соч.— 2-е изд.— Т. 14.— С. 276).

2.2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

Математическая модель линейной динамической системы может быть составлена на основе математических моделей элементов и звеньев, образующих систему. Линейная система в общем случае включает в себя звенья, соединенные последовательно, параллельно, охваченные обратными и перекрестными обратными связями. Передаточные функции всех этих структур могут выражаться через передаточные функции типовых структурных звеньев (рис. 2.2).

Последовательное соединение звеньев. В этом соединении выходная величина предыдущего звена является входной величиной последующего звена (рис. 2.2, а). Передаточная функция последовательно соединенных звеньев равна произведению передаточных функций всех звеньев, образующих соединение:

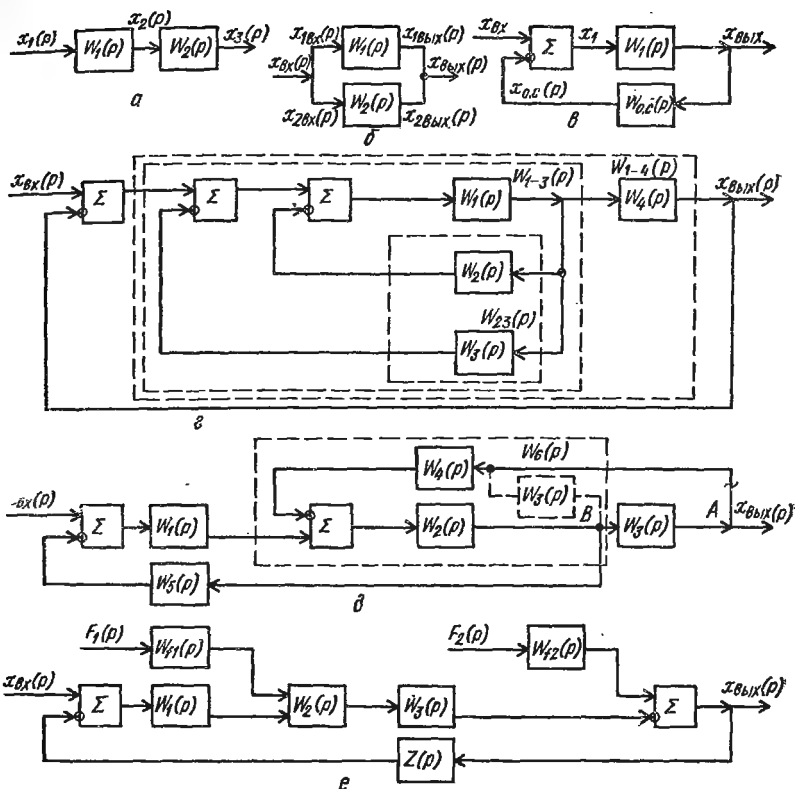


Рис. 2.2. Передаточные функции линейных динамических систем

$$W_1(p) = \frac{x_2(p)}{x_1(p)}; \quad W_2(p) = \frac{x_3(p)}{x_2(p)};$$

$$W(p) = \frac{x_3(p)}{x_1(p)} = \frac{W_2(p)x_2(p)W_1(p)}{x_2(p)} = W_1(p)W_2(p).$$

В общем случае

$$W(p) = \prod_{i=1}^n W_i(p), \quad i = 1, n, \quad (2.29)$$

где n — число последовательно соединенных звеньев.

Параллельное соединение звеньев. В этом соединении (рис. 2.2, б) на вход всех звеньев подается одна и та же величина, а выходная величина равна сумме выходных величин отдельных звеньев. На основании рис. 2.2, б имеем:

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{x_{1\text{вых}}(p) + x_{2\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{W_1(p)x_{\text{вх}}(p) + W_2(p)x_{\text{вх}}(p)}{x_{\text{вх}}(p)}$$

Так как $x_{1\text{вх}} = x_{2\text{вх}} = x_{\text{вх}}$, то $W(p) = W_1(p) + W_2(p)$ или в общем случае при k параллельно соединенных звеньях

$$W(p) = \sum_{i=1}^k W_i(p), \quad i = 1, k. \quad (2.30)$$

Таким образом, передаточная функция соединения из параллельных звеньев равна сумме их передаточных функций.

Звено, охваченное обратной связью. Для этого соединения (рис. 2.2, в)

$$x_1(p) = x_{\text{вх}}(p) \pm x_{\text{о.с}}(p),$$

где знак минус — для случая отрицательной обратной связи, плюс — для положительной.

Передаточная функция соединения

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{W_1(p) [x_{\text{вх}}(p) \pm W_{\text{о.с}}(p) x_{\text{вых}}(p)]}{x_{\text{вх}}(p)}$$

или после преобразования

$$W(p) = \frac{W_1(p)}{1 \pm W_1(p) W_{\text{о.с}}(p)}, \quad (2.31)$$

где знак плюс соответствует отрицательной обратной связи, а знак минус — положительной.

Общий случай соединения звеньев. Передаточная функция для общей схемы соединений звеньев (рис. 2.2, г) находится путем последовательного «сворачивания» отдельных частей схемы с использованием рассмотренных выше приемов нахождения эквивалентных передаточных функций. В результате для этого соединения имеем:

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{W_1(p) W_4(p)}{1 + W_1(p) [W_2(p) + W_3(p) + W_4(p)]}.$$

Соединения звеньев с перекрестными обратными связями. Для этих соединений характерно наличие звеньев, охваченных непараллельными обратными связями. На рис. 2.2, д звено с передаточной функцией $W_2(p)$ охвачено непараллельно соединенными звеньями с передаточными функциями $W_4(p)$ и $W_5(p)$. Непосредственное применение сформулированных выше правил для нахождения эквивалентной передаточной функции системы в рассматриваемом случае невозможно. Однако после некоторых преобразований эти правила становятся применимыми. Преобразования не должны нарушать динамических свойств системы, с одной стороны, а с другой — должны обеспечить возможность нахождения эквивалентной передаточной функции на основе указанных правил.

Приведенное на рис. 2.2, д преобразование отвечает поставленному условию. Преобразование свелось к тому, что обратная связь в точке А разомкнута и включена в точке В. В новую связь включено звено с передаточной функцией $W_3(p)$, что обеспечивает формирование на входе звена $W_4(p)$ такого же сигнала, как и в исход-

ной системе. Эквивалентную передаточную функцию $W_6(p)$ части системы, показанной штриховой линией, можно определить так:

$$W_6(p) = \frac{W_2(p)}{1 + W_2(p) W_3(p) W_4(p)}$$

Далее легко определяется передаточная функция всей системы.

Передаточная функция замкнутой системы. В замкнутой системе передаточная функция определяется как отношение выходной величины, преобразованной по Лапласу, к функции возмущающего воздействия, преобразованной по Лапласу при нулевых начальных условиях. Возмущающее воздействие может быть приложено к любой точке системы. Укажем общий способ нахождения передаточной функции по возмущению, независимый от точки приложения возмущения в системе (рис. 2.2, е).

Для определения передаточной функции по возмущению $F_1(p)$ положим $x_{\text{вх}}(p) = 0$ и $F_2(p) = 0$. Тогда

$$x_{\text{вых}}(p) = W_3(p) W_2(p) [F_1(p) W_{f1}(p) - W_1(p) z(p) x_{\text{вых}}(p)],$$

откуда

$$\frac{x_{\text{вых}}^{(F_1)}(p)}{F_1(p)} = \frac{W_{f1}(p) W_2(p) W_3(p)}{1 + W(p)}$$

где $W(p) = W_1(p) W_2(p) W_3(p) z(p)$ — передаточная функция разомкнутой системы.

Рассуждая аналогично, находим

$$\frac{x_{\text{вых}}^{(F_2)}(p)}{F_2(p)} = \frac{W_{f2}(p)}{1 + W(p)}; \quad \frac{x_{\text{вых}}^{(x_{\text{вх}})}(p)}{x_{\text{вх}}(p)} = \frac{W_1(p) W_2(p) W_3(p)}{1 + W(p)}$$

Полученные выражения показывают, что передаточная функция замкнутой системы по какому-либо воздействию представляет собой дробь, числитель которой равен произведению передаточных функций звеньев, включенных между выходной величиной и точкой приложения воздействия, а знаменатель — увеличенная на единицу передаточная функция разомкнутой системы.

Реакция системы на все воздействия определится как сумма реакций от каждого из воздействий:

$$x_{\text{вых}}(p) = x_{\text{вых}}^{(F_1)}(p) + x_{\text{вых}}^{(F_2)}(p) + x_{\text{вых}}^{(x_{\text{вх}})}(p),$$

т. е.

$$x_{\text{вых}}(p) = \frac{W_1(p) W_2(p) W_3(p) x_{\text{вх}}(p) + W_{f1}(p) W_2(p) W_3(p) F_1(p) + W_{f2}(p) F_2(p)}{1 + W(p)}$$

Отметим некоторые общие свойства передаточных функций линейных динамических систем, используемые при разработке и исследовании систем.

Передаточная функция любой линейной динамической системы может быть приведена к виду

$$W(p) = \frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{b_0 p^m + b_1 p^{m-1} + \dots + b_{m-1} p + b_m}{a_0 p^n + a_1 p^{n-1} + \dots + a_{n-1} p + a_n} \cdot \frac{K(p)}{H(p)} \quad (2.32)$$

где a и b — постоянные коэффициенты; $K(p) = b_0 p^m + b_1 p^{m-1} + \dots + b_{m-1} p + b_m$ — полином числителя передаточной функции системы; $H(p) = a_0 p^n + a_1 p^{n-1} + \dots + a_{n-1} p + a_n$ — полином знаменателя передаточной функции системы.

В частных случаях полином $K(p)$ может не содержать членов с оператором p и представлять собой произведение коэффициентов передачи звеньев, образующих систему. Например, в системе, состоящей только из колебательного звена,

$$\begin{aligned} K(p) &= k, \quad \text{т. е. } b_{m-1} = 0, \quad b_m = k; \\ H(p) &= T^2 p^2 + 2\xi T p + 1, \\ \text{т. е. } a_n &= 1, \quad a_{n-1} = 2\xi T, \quad a_{n-2} = T^2, \quad a_{n-3} = 0. \end{aligned}$$

В системе, состоящей из последовательного соединения усиленного звена, дифференцирующего звена первого порядка и колебательного звена, в соответствии с (2.29) имеем:

$$W(p) = \frac{K(\tau p + 1)}{T^2 p^2 + 2\xi T p + 1} = \frac{K(p)}{H(p)},$$

где

$$\begin{aligned} K(p) &= K(\tau p + 1); \quad b_m = K; \quad b_{m-1} = K\tau; \quad b_{m-2} = 0; \\ H(p) &= T^2 p^2 + 2\xi T p + 1; \quad a_n = 1; \quad a_{n-1} = 2\xi T; \\ &a_{n-2} = T^2; \quad a_{n-3} = 0. \end{aligned}$$

Важнейшим свойством выражения (2.32) является условие $n \geq m$, т. е. порядок полинома $H(p)$ знаменателя передаточной функции выше или не ниже порядка полинома $K(p)$ ее числителя. Это условие вытекает из физических свойств звеньев реальных динамических систем и более глубоко будет обосновано при рассмотрении частотных характеристик линейных динамических систем в следующем параграфе.

В отдельных случаях при идеализации динамических процессов в некоторых звеньях условие $n \geq m$ может быть нарушено. Например, для передаточной функции дифференцирующего звена первого порядка имеем $W(p) = \tau p + 1$; $K(p) = \tau p + 1$; $H(p) = 0$; $m = 1$; $n = 0$. Это обстоятельство лишь подтверждает условность и приближенность описания динамических режимов в системах управления с помощью линейных дифференциальных уравнений и передаточных функций.

Важно подчеркнуть, что коэффициенты a_i и b_i передаточной функции являются вещественными числами, так как они выражаются через конкретные физические параметры элементов системы.

Из выражения (2.32) передаточной функции системы можно получить дифференциальное уравнение системы в целом как в разомкнутом, так и в замкнутом состоянии.

Уравнения разомкнутых систем. Пусть выражение (2.32) является передаточной функцией разомкнутой системы. Тогда выражение

$$x_{\text{вых}}(p) H(p) = x_{\text{вх}}(p) K(p) \quad (2.33)$$

будет представлять собой операторное уравнение разомкнутой системы (уравнение в изображениях переменных). Положив в (2.33) $x_{\text{вх}}(p) = 0$, получим операторное уравнение свободного движения в разомкнутой линейной динамической системе

$$x_{\text{вых}}(p)H(p) = 0. \quad (2.34)$$

Переходя в (2.34) к оригиналам, т. е. переходя от операторного уравнения к дифференциальному и обозначив $x_{\text{вых}} = x$, получаем дифференциальное уравнение свободного движения в разомкнутой линейной динамической системе

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_{n-1} \frac{dx}{dt} + a_0 = 0. \quad (2.35)$$

Характеристическим уравнением, соответствующим дифференциальному уравнению (2.35), будет

$$H(p) = 0 \text{ или } a_0 p^n + a_1 p^{n-1} + \dots + a_{n-1} p + a_0 = 0. \quad (2.36)$$

Целесообразно подчеркнуть, что если в (2.34) символ p означает комплексную переменную или символ дифференцирования при нулевых начальных условиях, то в уравнении (2.36) символ p означает корень характеристического уравнения, соответствующего дифференциальному уравнению свободного движения в системе.

Из приведенных рассуждений следует весьма важный вывод: приравненный нулю знаменатель передаточной функции разомкнутой линейной динамической системы является характеристическим уравнением, соответствующим дифференциальному уравнению разомкнутой системы. В связи с этим многочлен $H(p) = 0$ называется *характеристическим оператором системы*.

Уравнение замкнутых систем. Пусть (2.32) является передаточной функцией разомкнутой системы. Для замкнутой системы в силу отрицательности главной обратной связи имеем $x_{\text{вх}} = -x_{\text{вых}}$, и (2.32) принимает вид $H(p)x_{\text{вых}}(p) = -K(p)x_{\text{вых}}(p)$. Обозначая $x_{\text{вых}} = x$, получим операторное уравнение свободного движения в замкнутой системе

$$[K(p) + H(p)]x(p) = 0, \quad (2.37)$$

где $K(p)$, $H(p)$ — соответственно числитель и знаменатель передаточной функции разомкнутой системы; $x(p)$ — изображение координаты системы в точке ее замыкания.

На основании (2.37) можно записать характеристическое уравнение, соответствующее дифференциальному уравнению свободного движения в замкнутой системе

$$K(p) + H(p) = 0. \quad (2.38)$$

По уравнению (2.31), с учетом того, что $W_{0,c}(p) = 1$, передаточная функция замкнутой системы

$$W_3(p) = W(p)/[1 + W(p)], \quad (2.39)$$

где $W(p) = K(p)/H(p)$ — передаточная функция разомкнутой системы.

Преобразуя (2.39), найдем

$$W_3(p) = \frac{K(p)}{K(p) + H(p)} = \frac{K(p)}{H_3(p)} \quad (2.40)$$

На основании (2.38) и (2.40) характеристическое уравнение замкнутой системы можно записать в виде

$$H_3(p) = 0. \quad (2.41)$$

Таким образом, из (2.38) и (2.41) следует, что приравненная нулю сумма полинома числителя и полинома знаменателя передаточной функции разомкнутой системы или приравненный нулю полином знаменателя передаточной функции замкнутой системы являются характеристическим уравнением, соответствующим дифференциальному уравнению свободного движения в замкнутой системе.

Уравнение вынужденного движения в замкнутой системе, т. е. уравнение переходного процесса под влиянием возмущающего воздействия, может быть получено из передаточной функции замкнутой системы по возмущающему воздействию. Способ определения передаточной функции по возмущающему воздействию, приложенному в любой точке системы, рассмотрен выше (см. рис. 2.2, е).

2.3. ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

В практике анализа и синтеза линейных динамических систем управления большую роль играют частотные характеристики, отражающие свойства систем, работающих в режиме гармонических воздействий. К таким характеристикам относятся амплитудно-фазовые (АФХ) и логарифмические амплитудно-фазовые частотные характеристики (ЛАФЧХ).

Аналитическое выражение АФХ формально получается из выражения передаточной функции при чисто мнимом значении ее аргумента: $p = j\omega$. Таким образом, АФХ, т. е. $W(j\omega)$, отражает отношение выходной величины ко входной при гармоническом воздействии на входе.

Пусть к системе с передаточной функцией $W(p)$ приложено гармоническое входное воздействие с частотой ω :

$$x_{\text{вх}} = X_{\text{вх}} e^{j\omega t}, \quad (2.42)$$

где $X_{\text{вх}}$ — амплитуда входного воздействия.

Если система устойчива, то собственные колебания с течением времени исчезнут и останутся лишь вынужденные движения. При этом в линейной системе на выходе установятся также гармонические колебания с той же частотой ω :

$$x_{\text{вых}} = X_{\text{вых}} e^{j(\omega t + \varphi)}, \quad (2.43)$$

где $X_{\text{вых}}$ — амплитуда выходных колебаний; φ — угол сдвига фаз между выходными и входными колебаниями.

Взяв отношение выходной величины к входной, на основании (2.43) и (2.42) имеем:

$$\frac{X_{\text{вых}} e^{j(\omega t + \varphi)}}{X_{\text{вх}} e^{j\omega t}} = A(\omega) e^{j\varphi(\omega)} = W(j\omega), \quad (2.44)$$

где $A(\omega) = X_{\text{вых}}/X_{\text{вх}}$ — амплитуда вектора $W(j\omega)$.

Для каждого значения частоты ω возмущающего воздействия на входе выражение $W(j\omega)$ в комплексной плоскости представляется вектором, амплитуда $A(\omega)$ и фаза $\varphi(\omega)$ которого являются функциями частоты. С изменением частоты в диапазоне $-\infty < \omega < +\infty$ конец вектора $W(j\omega)$ будет скользить по некоторой кривой, расположенной в комплексной плоскости.

Геометрическое место точек конца вектора $W(j\omega)$ на комплексной плоскости при изменении частоты в указанном диапазоне называется *амплитудно-фазовой характеристикой* (АФХ) динамической системы с передаточной функцией $W(p)$.

Ищем общий прием построения АФХ и проиллюстрируем его на конкретном примере. АФХ любого линейного звена или системы строится на основе их передаточных функций. Заменяя $p = j\omega$, получим исходное выражение для АФХ:

$$W(j\omega) = \frac{K(j\omega)}{H(j\omega)} = A(\omega) e^{j\varphi(\omega)}. \quad (2.45)$$

Далее, в выражении (2.45) нужно избавиться от мнимости в знаменателе, для чего числитель и знаменатель его умножим на выражение, сопряженное знаменателю:

$$W(j\omega) = \frac{K(j\omega)}{H(j\omega)} = \frac{K(j\omega) \overline{H(j\omega)}}{[\text{mod} H(j\omega)]^2}, \quad (2.46)$$

где $\overline{H(j\omega)}$ — выражение, сопряженное выражению $H(j\omega)$.

Отделяя в (2.46) вещественную часть от мнимой, получим

$$W(j\omega) = P(\omega) + jQ(\omega), \quad (2.47)$$

где $P(\omega)$ — полином, содержащий четные степени ω , т. е. вещественная часть выражения $W(j\omega)$; $Q(\omega)$ — полином, содержащий нечетные степени ω , т. е. мнимая часть выражения $W(j\omega)$.

Нетрудно видеть, что $P(\omega)$ является четной функцией аргумента ω , а $Q(\omega)$ — нечетной, что свидетельствует о симметричности кривой $W(j\omega)$ относительно вещественной оси. Поэтому, построив АФХ в диапазоне частот $0 < \omega < +\infty$, вторую часть АФХ, т. е. часть в диапазоне $-\infty < \omega < 0$, получим как зеркальное отображение первой ее части относительно вещественной оси.

Рассмотрим построение АФХ на примере колебательного звена.

В выражении передаточной функции колебательного звена (2.12) при коэффициенте усиления $k=1$

$$W(p) = \frac{1}{T^2 p^2 + 2\xi T p + 1}$$

положим $p = j\omega$ и, принимая во внимание, что $j^2 = -1$, получим

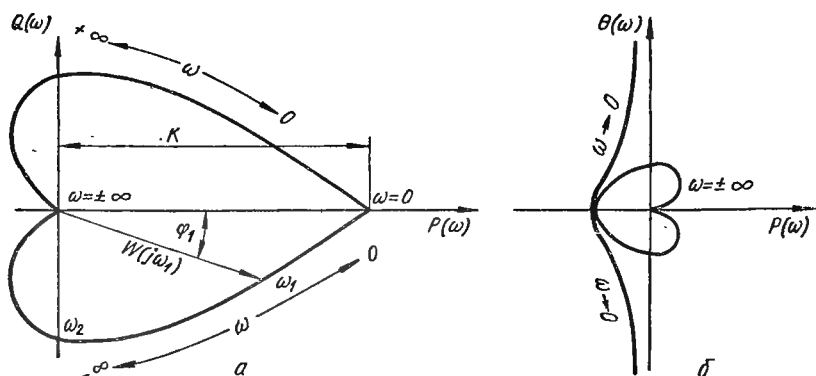


Рис. 2.3. Амплитудно-фазовые характеристики:
а — колебательного звена; б — астатической системы

$$W(j\omega) = \frac{1}{(1 - T^2\omega^2) + j2\xi\omega T}$$

Умножая числитель и знаменатель последнего уравнения на выражение, сопряженное знаменателю, и отделяя вещественную часть от мнимой, получим

$$\begin{aligned} W(j\omega) &= \frac{1 - T^2\omega^2}{(1 - T^2\omega^2)^2 + (2\xi\omega T)^2} - j \frac{2\xi\omega T}{(1 - T^2\omega^2)^2 + (2\xi\omega T)^2} = \\ &= P(\omega) + jQ(\omega), \end{aligned} \quad (2.47)$$

где $P(\omega) = \frac{1 - T^2\omega^2}{(1 - T^2\omega^2)^2 + (2\xi\omega T)^2}$ — вещественная часть; $Q(\omega) = \frac{2\xi\omega T}{(1 - T^2\omega^2)^2 + (2\xi\omega T)^2}$ — коэффициент при мнимой части.

Построенная по уравнению (2.47) АФХ колебательного звена приведена на рис. 2.3, а. Из этого рисунка видно, что для любого звена

$$\begin{aligned} A(\omega) &= \sqrt{[P(\omega)]^2 + [Q(\omega)]^2}, \\ \varphi(\omega) &= \arctg \frac{Q(\omega)}{P(\omega)}. \end{aligned} \quad (2.48)$$

Учитывая (2.48) и (2.57), для колебательного звена находим:

$$\begin{aligned} A(\omega) &= \frac{k}{\sqrt{(1 - T^2\omega^2)^2 + (2\xi\omega T)^2}} = \frac{X_{\text{ВЫХ}}}{X_{\text{ВХ}}}; \\ \varphi(\omega) &= -\arctg \frac{2\xi T\omega}{1 - T^2\omega^2}. \end{aligned}$$

Модуль АФХ представляет собой отношение амплитуды колебаний выходной величины звена или системы к амплитуде гармонических колебаний с частотой ω на входе, а аргумент вектора

АФХ равен углу сдвига фаз между выходной и входной величинами.

Уравнения (2.45) и (2.48) являются исходными для построения АФХ любой системы с передаточной функцией $W(p)$. Величины $P(\omega)$ и $Q(\omega)$ определяются в порядке, описанном выше.

Отметим особенность АФХ астатических систем (рис. 2.3, б), вытекающую из наличия в этих системах интегрирующих звеньев, вследствие чего передаточная функция таких систем имеет вид

$$W(p) = K(p)/[p^\nu H(p)], \quad (2.49)$$

где ν — порядок астатизма.

Из (2.49) видно, что при $p=j\omega$ и $\omega \rightarrow 0$ модуль вектора АФХ астатической системы стремится к бесконечности.

Логарифмической амплитудно-частотной характеристикой (ЛАЧХ) звена или системы называется построенная в логарифмическом масштабе частот функция вида

$$L(\omega) = 20 \lg A(\omega), \quad (2.50)$$

где $A(\omega)$ — модуль АФХ звена или системы, определяемый согласно выражению (2.48). Функция $L(\omega)$ измеряется в децибеллах (дБ). С учетом физического смысла модуля $A(\omega)$ следует, что 1 дБ представляет собой 20 десятичных логарифмов отношения амплитуды выходной величины звена или системы к амплитуде входной величины.

Функция $\varphi(\omega)$, определяемая выражением (2.48) и построенная в логарифмическом масштабе частот, называется логарифмической фазочастотной характеристикой (ЛФЧХ).

Построенные на одном графике характеристики $L(\omega)$ и $\varphi(\omega)$ представляют собой ЛАФЧХ.

Использование ЛАФЧХ позволяет легко и быстро оценить динамические свойства автоматических систем и выбрать необходимые дополнительные звенья, введение которых в систему обеспечит требуемые свойства. Упрощенными методами можно без каких-либо затруднений построить эти характеристики как для отдельных звеньев, так и для системы в целом.

Непосредственное использование выражений (2.48) и (2.50) для построения ЛАФЧХ приводит к громоздким и трудоемким расчетам. Однако при вполне приемлемых допущениях можно существенно сократить объем работы по построению ЛАФЧХ с достаточной для инженерной практики точностью. Рассмотрим приближенные методы построения ЛАФЧХ типовых структурных звеньев и линейных динамических систем, состоящих из этих звеньев.

Апериодическое звено

Используя выражение передаточной функции аperiодического звена (2.3), уравнения (2.48) и изложенный выше порядок построения АФХ, найдем при $k=1$

$$A(\omega) = \frac{1}{\sqrt{1 + \omega^2 T^2}}; \quad \varphi(\omega) = -\arctg(\omega T). \quad (2.51)$$

Далее, на основании (2.50) получим выражение ЛАФЧХ апериодического звена:

$$L(\omega) = -20 \lg \sqrt{1 + \omega^2 T^2}. \quad (2.52)$$

Условие $k=1$, не имея принципиального значения при описании порядка построения ЛАФЧХ, упрощает рассуждения.

Из анализа (2.52) видно, что при $\omega \ll 1/T$ выражение $\omega^2 T^2 \approx 0$, следовательно, $L(\omega) = 0$. При $\omega \gg 1/T$ функция $L(\omega) = -20 \lg \omega T$. Для некоторой частоты $\omega_1 \gg 1/T$ имеем $L(\omega_1) = -20 \lg \omega_1 T$.

Увеличим частоту в 10 раз, т. е. примем $\omega_2 = 10 \omega_1$ (заметим, что увеличение частоты в 10 раз называется *возрастанием частоты на декаду*). Тогда $L(\omega_2) = -20 \lg 10 \omega_1 T = -20 \lg \omega_1 T - 20 = L(\omega_1) - 20$. Рассуждая аналогично, для $\omega_3 = 100 \omega_1$ получим $L(\omega_3) = L(\omega_1) - 40$.

Таким образом, при увеличении частоты на декаду характеристика $L(\omega)$ апериодического звена убывает на 20 дБ. В результате порядок построения ЛАФЧХ апериодического звена сводится к следующему: левее частоты $\omega = 1/T$ характеристика $L(\omega)$ совпадает с осью абсцисс, а правее этой частоты $L(\omega)$ представляет собой прямую линию с наклоном — 20 дБ/дек. (рис. 2.4, а). Частота $\omega = 1/T = \omega_c$, при которой характеристика $L(\omega)$ изменяет свой наклон, называется *сопрягающей частотой*. На рисунке по оси абсцисс отложены в логарифмическом масштабе значения частот в относительных единицах, причем за базовую частоту принята частота ω_c . Иначе говоря, на оси абсцисс в логарифмическом масштабе откладываются значения частот, поделенные на ω_c . Крайняя левая граница рисунка (положение оси ординат функции $L(\omega)$) определяется диапазоном частот, в котором рассматривается динамика звена.

Оценим максимальную погрешность, которая вносится при построении характеристики $L(\omega)$ апериодического звена описанным способом. Очевидно, что максимальная погрешность наступает при $\omega = \omega_c = 1/T$. В соответствии с (2.52) для этой частоты $L(\omega) = -3$ дБ, а в случае приближенного построения для этой частоты $L(\omega) = 0$. Насколько допустима такая ошибка, можно судить по результатам анализа каждой конкретной системы. При необходимости в приближенные расчеты функции $L(\omega)$ можно внести уточнения в районе частот, прилегающих к сопрягающей частоте.

Логарифмическая фазочастотная характеристика апериодического звена $\varphi(\omega)$ строится по выражению (2.51), из которого следует, что $\varphi(\omega) \rightarrow 0$ при $\omega \rightarrow 0$; $\varphi(\omega) \rightarrow (-90^\circ)$ при $\omega \rightarrow \infty$; $\varphi(\omega_c) = -\arctg 1 = -45^\circ$ при $\omega = \omega_c = 1/T$. Таким образом, характеристика апериодического звена представляет собой кососимметричную кривую относительно ординаты $\omega = \omega_c$ и абсциссы $\varphi(\omega) = -45^\circ$, изменяясь в диапазоне от 0 до -90° .

Построение характеристики $\varphi(\omega)$ существенно упрощается, если применять номограмму (рис. 2.4, б), на которой снизу в логарифмическом масштабе нанесены значения $\omega/\omega_c = \omega T$, а сверху — определенные по выражению (2.51) фазовые сдвиги в градусах, соответствующие значениям отношений ω/ω_c . Расположив нижнюю

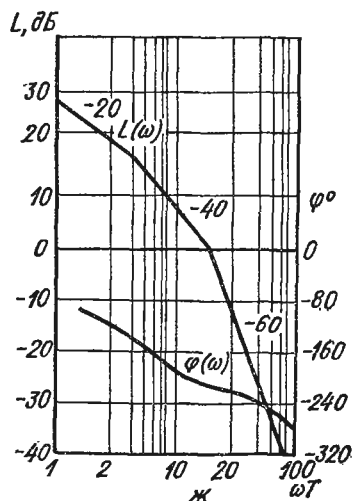
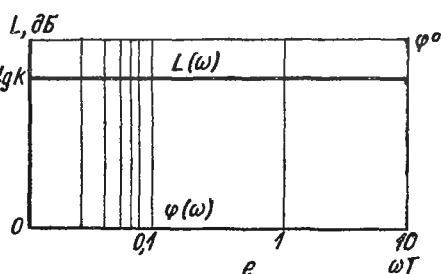
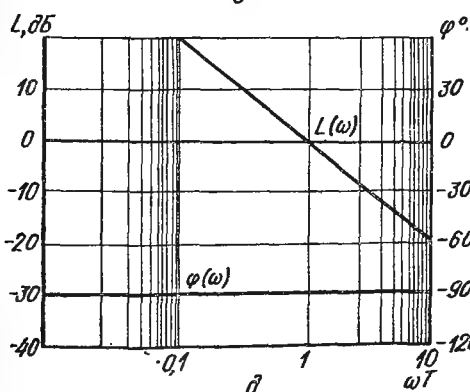
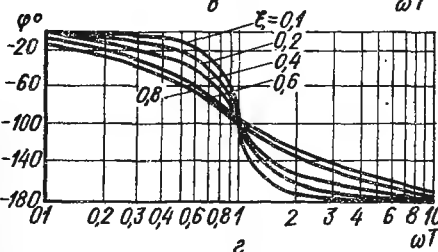
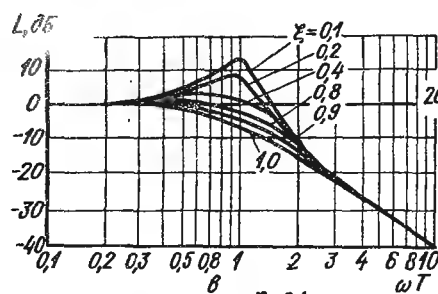
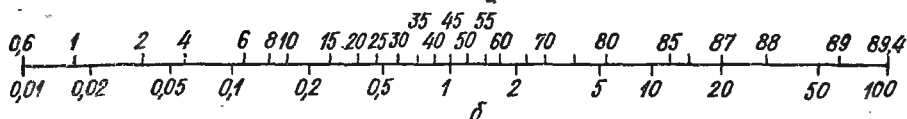
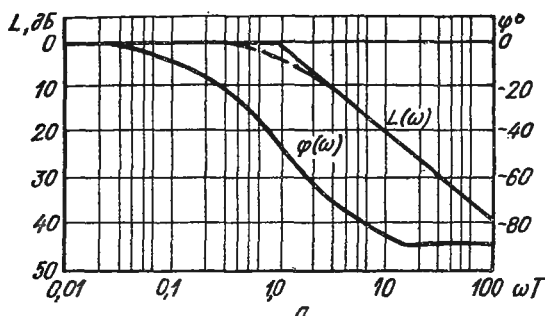


Рис. 2.4. Логарифмические амплитудно-фазовые частотные характеристики: а — аperiodического звена; б — номограмма для определения фазовой характеристики аperiodического звена; в — ЛАЧХ — колебательного звена; г — ЛФЧХ колебательного звена; д — ЛАФЧХ интегрирующего звена; е — усиливающего звена; ж — разомкнутой системы

шкалу номограммы на характеристике $L(\omega)$ так, чтобы отметка $\omega/\omega_c = 1$ совпала с частотой сопряжения, сдвиг фазы при любой другой частоте определяют по верхней шкале над этим значением частоты. Непременным условием применения номограммы является одинаковость масштабов номограммы и оси абсцисс характеристики $L(\omega)$. Построение ЛАФЧХ звеньев и систем существенно упрощается при выборе такого масштаба по оси частот, при котором одна декада соответствует десяти сантиметрам. В таком случае логарифмы частот легко определяются по нижней шкале логарифмической линейки.

Колебательное звено

Аналитическое выражение ЛАЧХ колебательного звена получим на основании уравнений (2.47), (2.48):

$$L(\omega) = -20 \lg \sqrt{(1 - T^2 \omega^2)^2 + (2\xi \omega T)^2}. \quad (2.53)$$

При $\omega \ll 1/T$ функция $L(\omega) \approx 0$; при $\omega \gg 1/T$ функция $L(\omega) \approx -20 \lg \omega^2 T^2$.

Проведя рассуждения, аналогичные тем, что имели место при определении формы $L(\omega)$ апериодического звена, получим следующую форму ЛАЧХ колебательного звена (рис. 2.4, в): в районе $\omega \ll 1/T$ характеристика $L(\omega)$ совпадает с осью абсцисс; в районе $\omega \gg 1/T$ она представляет собой прямую линию с наклоном — 40 дБ/дек. В отличие от апериодического звена, погрешность построения, характеристики $L(\omega)$ колебательного звена при $\omega = \omega_c = 1/T$ может быть значительной. Она не превышает 3 дБ при условии $0,38 \leq \xi \leq 0,7$. При других значениях коэффициента колебательности ξ характеристику $L(\omega)$ в районе частоты среза необходимо строить непосредственно по выражению (2.53) или пользоваться соответствующей таблицей поправок.

Фазовая характеристика колебательного звена строится на основании выражения, вытекающего из уравнений (2.47) и (2.48):

$$\varphi(\omega) = -\arctg \frac{Q(\omega)}{P(\omega)} = -\arctg \frac{2\xi T \omega}{1 - T^2 \omega^2}. \quad (2.54)$$

Характеристика $\varphi(\omega)$ (рис. 2.4, г) представляет собой кососимметричную кривую относительно осей $\omega = \omega_c$, $\varphi(\omega) = 90^\circ$.

Интегрирующее звено

Для получения выражения ЛАФЧХ интегрирующего звена в уравнении его передаточной функции $W(p) = 1/p$ положим $p = j\omega$, избавимся от мнимости в знаменателе и, используя выражения (2.48), найдем

$$A(\omega) = 1/\omega; \quad \varphi(\omega) = \arctg(-\infty) = -\pi/2. \quad (2.55)$$

На основании (2.50)

$$L(\omega) = 20 \lg A(\omega) = -20 \lg \omega.$$

Из этого выражения видно, что $L(\omega) = 20$ дБ при $\omega = 0,1$; $L(\omega) = 0$ при $\omega = 1$; $L(\omega) = -20$ дБ при $\omega = 10$.

Таким образом, ЛАЧХ интегрирующего звена представляет собой линию с наклоном -20 дБ/дек., пересекающую ось абсцисс в точке $\omega = 1$ (рис. 2.4, д). Как видно из (2.55), фазовая характеристика интегрирующего звена представляет собой прямую линию, параллельную оси абсцисс со значением угла $\varphi(\omega) = -90^\circ$, т. е. сдвиг фазы выходной величины по отношению к входной в интегрирующем звене не зависит от частоты и равен -90° .

Усилительное звено

В уравнении передаточной функции усилительного звена $W(p) = k$, положив $p = j\omega$, получим $W(j\omega) = k$; $P(\omega) = k$; $Q(\omega) = 0$, тогда

$$A(\omega) = \sqrt{P^2(\omega) + Q^2(\omega)} = k; \quad L(\omega) = 20 \lg(\omega) = 20 \lg k; \quad (2.56)$$

$$\varphi(\omega) = \operatorname{arctg} \frac{Q(\omega)}{P(\omega)} = 0^\circ; \quad (2.57)$$

Из (2.56) и (2.57) следует, что ЛАЧХ усилительного звена представляет собой прямую линию, проходящую на высоте $20 \lg k$ (рис. 2.4, е), а ЛФЧХ — прямая, совпадающая с осью абсцисс.

Дифференцирующее звено

Исходя из уравнений передаточных функций идеального дифференцирующего звена (2.22), дифференцирующего звена первого порядка (2.27) и второго порядка (2.28) и проведя рассуждения, аналогичные вышеприведенным, можно прийти к выводу, что ЛАФЧХ идеального дифференцирующего звена является зеркальным отображением ЛАФЧХ интегрирующего звена, ЛАФЧХ дифференцирующего звена первого порядка — аperiodического звена, а ЛАФЧХ дифференцирующего звена второго порядка — колебательного звена. Таким образом, например, ЛАЧХ дифференцирующего звена первого порядка до частоты $\omega = \omega_c/\tau$ совпадает с осью частот, а при $\omega = 1/\tau$ представляет собой прямую линию, проходящую с наклоном $+20$ дБ/дек. ЛФЧХ этого звена представляет собой линию, кососимметричную относительно оси ординат $\omega = 1/\tau$ и оси абсцисс $\varphi(\omega) = 45^\circ$.

Одноконтурные разомкнутые системы

Для построения ЛАФЧХ разомкнутой системы целесообразно систему представить в виде совокупности типовых структурных звеньев. Тогда характеристики $L(\omega)$ и $\varphi(\omega)$ системы можно получить, суммируя ординаты соответствующих характеристик, построенных для каждого типового звена, входящего в систему. На практике это построение еще более упрощается и не требует построения характеристик отдельных звеньев во всем диапазоне час-

тот. Рассмотрим характерный пример, раскрывающий приемы построения ЛАФЧХ одноконтурных линейных динамических систем.

Пример. Построим характеристики $L(\omega)$ и $\varphi(\omega)$ одноконтурной системы, имеющей передаточную функцию

$$W(p) = \frac{k}{(1+T_1p)(1+T_2p)p} = \frac{24}{(1+0,07p)(1+0,28p)p}. \quad (2.58)$$

Как видно из (2.58), система включает в себя усилительное звено с коэффициентом усиления $k = 24$, два апериодических звена с постоянными времени $T_1 = 0,07$ с и $T_2 = 0,28$ с и одно интегрирующее звено.

Для построения характеристики $L(\omega)$ системы найдем сопрягающие частоты:

$$\omega_{c1} = 1/T = 1/0,07 = 14 \text{ с}^{-1}; \quad \omega_{c2} = 1/T = 1/0,28 = 3,57 \text{ с}^{-1}.$$

Само построение удобно начинать с характеристик интегрирующих звеньев (или дифференцирующих идеальных, если такие имеются в системе). С этой целью на высоте $20 \lg k = 20 \lg 24 = 27,6$ дБ при $\omega = 1$ ($\lg \omega = 0$) проведем линию с наклоном -20 дБ/дек., отражающую ЛАЧХ выражения $k/p = 24/p$ (рис. 2.4, ж). Эта линия до пересечения с ординатой $\omega_{c2} = 3,57$, с^{-1} является характеристикой $L(\omega)$ всей системы, так как на этом участке частот логарифмические амплитудно-частотные характеристики апериодических звеньев совпадают с осью абсцисс и их ординаты равны нулю.

При частотах $\omega_{c2} < \omega < \omega_{c1}$ на характеристику $L(\omega)$, отражающую передаточную функцию $W(p) = 24/p$, накладывается ЛАЧХ апериодического звена с постоянной времени $T_2 = 0,28$ с и сопрягающей частоты $\omega_{c2} = 3,57 \text{ с}^{-1}$. Следовательно, на этом участке характеристика $L(\omega)$ системы будет иметь наклон -40 дБ/дек. Продолжая рассуждения, приходим к выводу, что при $\omega > \omega_{c1}$, т. е. при $\omega > 14,3 \text{ с}^{-1}$, характеристика $L(\omega)$ имеет наклон -60 дБ/дек.

Рассмотренный прием построения ЛАЧХ одноконтурной разомкнутой системы является универсальным и удобным для построения характеристики $L(\omega)$ любой одноконтурной системы, представленной в виде последовательного соединения типовых структурных звеньев. Как видно, этот прием сводится к тому, что сначала проводится участок ЛАЧХ, соответствующий интегрирующим (или идеальным дифференцирующим) звеньям с учетом коэффициента усиления, а далее наклон характеристики изменяется при каждой из сопрягающих частот, а именно: увеличивается на 20 дБ/дек. в апериодическом звене, увеличивается на 40 дБ/дек. в колебательном звене, уменьшается на 20 дБ/дек. в дифференцирующем звене первого порядка и уменьшается на 40 дБ/дек. в дифференцирующем звене второго порядка.

Для построения фазовых характеристик одноконтурных систем удобно пользоваться номограммой (рис. 2.4, б). По ней определяются значения фазовых сдвигов для каждого звена, а фазовые сдвиги системы находятся как сумма значений фазовых углов отдельных звеньев. Хотя номограмма построена для

Таблица 2.2

Фазовая характеристика	Значение фазы, град, при частоте ω , с^{-1}					
	2	5	10	20	30	50
$\varphi_1(\omega)$	$\varphi_1(\omega) = 0$					
$\varphi_2(\omega)$	-8	-19,3	-35	-54,3	-64,1	-74,1
$\varphi_3(\omega)$	-29,3	-54,5	-70,5	-79,8	-83,1	-88,0
$\varphi_4(\omega)$	$\varphi_4(\omega) = -90^\circ$					
$\varphi(\omega)$	-127,3	-163,8	-195,5	-224,1	-237,2	-270,1

определения фазовых сдвигов в апериодическом звене, она может использоваться для построения характеристики $\varphi(\omega)$ любого звена. При этом для колебательного звена значения фазовых сдвигов, полученных из номограммы, нужно удвоить. Фазовый угол для дифференцирующего звена первого порядка будет по величине такой же, как и для апериодического звена, но с обратным знаком (плюс), а для дифференцирующего звена второго порядка такой же, как и для колебательного звена, но с обратным знаком (плюс). Напомним, что в усилительных, интегрирующих и идеальных дифференцирующих звеньях фазовые сдвиги имеют постоянные значения и не зависят от частоты: в усилительном звене $\varphi(\omega) = 0$; в интегрирующем $\varphi(\omega) = -90^\circ$; в идеальном дифференцирующем $\varphi(\omega) = 90^\circ$.

В табл. 2.2 занесены числовые значения фазовых сдвигов в рассматриваемой одноконтурной системе, определенные с помощью номограммы, а результирующая характеристика $\varphi(\omega)$ системы приведена на рис. 2.4, *ж*.

Замкнутые системы

Динамические свойства замкнутых линейных систем в значительной мере определяются с помощью ЛАФЧХ разомкнутых систем. Тем не менее при анализе замкнутых многоконтурных систем возникает необходимость построения логарифмических амплитудно-частотной $L_3(\omega)$ и фазочастотной $\varphi_3(\omega)$ характеристик.

Характеристики $L_3(\omega)$ и $\varphi_3(\omega)$ могут быть определены по соответствующим характеристикам $L(\omega)$ и $\varphi(\omega)$ разомкнутого контура. Действительно, если разомкнутая система имеет передаточную функцию $W(p)$, то соответствующая замкнутая система будет иметь передаточную функцию

$$W_3(p) = W(p) / [1 + W(p)],$$

а выражение для амплитудно-фазовой характеристики замкнутой системы имеет вид

$$A_3(\omega) e^{j\varphi_3(\omega)} = \frac{A(\omega) e^{j\varphi(\omega)}}{1 + A(\omega) e^{j\varphi(\omega)}}$$

Из последнего уравнения, используя свойства показательных функций, можно получить

$$A(\omega) = \frac{1}{\sqrt{1 - 2[A_3(\omega)]^{-1} \cos \varphi_3(\omega) + [A_3(\omega)]^{-2}}};$$

$$\varphi(\omega) = \arctg \frac{\sin \varphi_3(\omega)}{\cos \varphi_3(\omega) - A_3(\omega)}$$

На основании этих выражений могут быть построены так называемые номограммы замыканий (рис. 2.5), с помощью которых легко найти характеристики $L_3(\omega)$ и $\varphi_3(\omega)$ по характеристикам $L(\omega)$ и $\varphi(\omega)$ разомкнутой системы. На номограмме в координатных осях $L(\omega) = 20 \lg A(\omega)$ и $\varphi(\omega)$ нанесены линии равных значений $A_3(\omega)$ при переменном значении $\varphi_3(\omega)$ и линии равных значений $\varphi_3(\omega)$ при переменном значении $A_3(\omega)$. На кривых приведены значения фазы и модуля. Зная значения $L(\omega)$ и $\varphi(\omega)$ при заданном значении частоты и нанеся точку с этими координатами на номограмму, можно найти значения $L_3(\omega)$ и $\varphi_3(\omega)$.

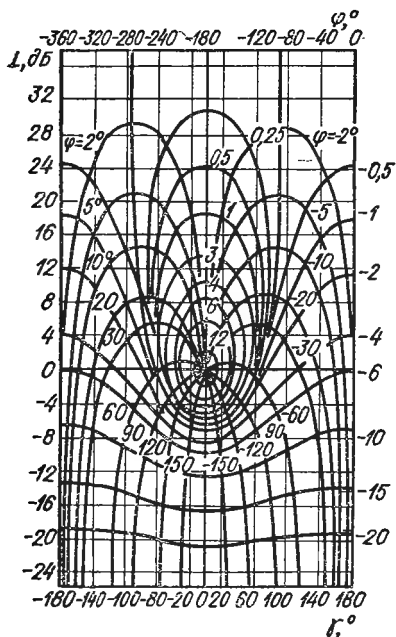


Рис. 2.5. Номограммы замыкания (γ° — избыток фазы)

определить по отношению коэффициентов передаточной функции. Не останавливаясь на обосновании этого метода построения ЛАЧХ, укажем порядок построения характеристик $L(\omega)$ и $\varphi(\omega)$ на базе отношений коэффициентов передаточной функции многоконтурной системы.

Для использования коэффициентного метода передаточную функцию статической системы целесообразно представить в виде

$$W(p) = \frac{K(p)}{K(p)} k = k \frac{b_m p^m + b_{m-1} p^{m-1} + \dots + b_1 p + 1}{a_n p^n + a_{n-1} p^{n-1} + \dots + a_1 p + 1}, \quad (2.59)$$

а передаточную функцию астатической системы — в виде

$$W(p) = \frac{k}{p^v} \frac{K(p)}{H(p)} = \frac{k}{p^v} \frac{b_m p^m + b_{m-1} p^{m-1} + \dots + b_1 p + 1}{a_n p^n + a_{n-1} p^{n-1} + \dots + a_1 p + 1} \quad (2.60)$$

где v — порядок астатизма.

Введем определения. Сопрягающие частоты числителя $\omega_{\text{ч}}$ и знаменателя $\omega_{\text{з}}$ выразим так:

$$\omega_{\text{ч}1} = 1/b_1; \quad \omega_{\text{ч}2} = b_1/b_2; \quad \dots; \quad \omega_{\text{ч}m} = b_{m-1}/b_m; \\ \omega_{\text{з}1} = 1/a_1; \quad \omega_{\text{з}2} = a_1/a_2; \quad \dots; \quad \omega_{\text{з}n} = a_{n-1}/a_n.$$

С учетом принятых допущений и определений ЛАЧХ статической системы строится следующим образом. В координатной системе $L(\omega) - \lg \omega$ на оси частот отложим значения $\omega_{\text{ч}}$ и $\omega_{\text{з}}$ в порядке их возрастания. Далее, на высоте $20 \lg k$ проведем прямую, парал-

ЛАЧХ звена охваченного обратной связью, приближенно можно выразить через ЛАЧХ обратной связи. Действительно, АФХ такого соединения имеет вид

$$W(j\omega) = \frac{W_1(j\omega)}{1 + W_1(j\omega)W_{o.c}(j\omega)},$$

где $W_1(j\omega)$ — АФХ прямого звена; $W_{o.c}(j\omega)$ — АФХ звена обратной связи.

Если $|W_1(j\omega)W_{o.c}(j\omega)| \gg 1$, то $W(j\omega) \approx 1/[W_{o.c}(j\omega)]$, следовательно, для этих условий

$$L(\omega) = -L_{o.c}(\omega),$$

т. е. ЛАЧХ звена, охваченного обратной связью, может быть замещена ЛАЧХ обратной связи, взятой с обратным знаком.

В тех случаях, когда постоянные времени структурных звеньев в системе существенно отличаются, например на порядок, сопрягающие частоты можно приближенно опре-

лельную оси абсцисс до пересечения с ординатой первой сопрягающей частоты. Если эта частота определена как ω_3 , то наклон характеристики $L(\omega)$ увеличивается на 20 дБ/дек., если частота определена как ω_4 , то наклон уменьшается на 20 дБ/дек., что отражает влияние звеньев с отставанием выходного сигнала по фазе (аперриодических и колебательных) и звеньев с положительным сдвигом выходного сигнала (дифференцирующих).

Для астатических систем порядок построения $L(\omega)$ отличается тем, что первый участок характеристики проводится через точку с координатами $20 \lg k$, $\omega = 1$ с наклоном — 20 в дБ/дек. Далее ЛАЧХ строится так же, как и для статических систем.

Для построения фазовой характеристики можно использовать номограмму (рис. 2.4, а) при условии, что фазовые сдвиги, определяемые сопрягающими частотами знаменателя, берутся с отрицательными знаками, а фазовые сдвиги, определяемые сопрягающими частотами числителя, — с положительными знаками. При определении суммарных значений $\varphi(\omega)$ необходимо учитывать фазовые сдвиги интегрирующих (-90°) и идеальных дифференцирующих ($+90^\circ$) звеньев, если таковые имеются в системе.

Использование ЛАФЧХ для анализа линейных динамических систем рассмотрено в последующих главах.

2.4. УСТОЙЧИВОСТЬ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

Важнейшей задачей анализа динамических систем является решение вопроса об их устойчивости, т. е. наличия или отсутствия в системе способности возвращаться к равновесному состоянию, из которого она выводится возмущающими или управляющими воздействиями. Ранее отмечалось, что неустойчивая система неработоспособна, так как любое воздействие на такую систему вызывает недопустимое возрастание физических величин, характеризующих работу системы (токи, электрические напряжения, ускорения, механические напряжения и т. д.).

Возникновение неустойчивых (расходящихся) колебаний в системе можно объяснить, например, проследив включение системы управления величиной $x_{\text{вых}}$ (см. рис. 1.2, б). Управляющая величина на входе регулятора x_p в режиме установившейся работы системы определяется как разность $x_p = x_{\text{вх}} - x_{0.c}$. В первый момент включения системы в силу инерционности обратной связи $x_{0.c} = 0$, следовательно, $x_p = x_{\text{вх}}$, т. е. $x_p \gg x_{p.\text{уст}}$, где $x_{p.\text{уст}}$ — управляющее воздействие на входе регулятора. Естественно, что это может вызвать значительное возрастание выходной величины $x_{\text{вых}} \gg x_{\text{вых.уст}}$, которая будет возрастать до тех пор, пока сигнал обратной связи не уменьшит значения x_p . Однако значительно возросшая величина $x_{\text{вых}}$ через ОС передается на вход системы и существенно уменьшит значение x_p , что может привести к последующему снижению величины $x_{\text{вых}}$ и возникновению условия $x_{\text{вых}} \ll x_{\text{вых.уст}}$. При неблагоприятном соотношении параметров системы описанные колебательные процессы могут стать расходящимися.

Задача исследования систем на устойчивость может быть поставлена двояко: 1) устойчива ли система при заданном значении ее параметров; 2) в каких диапазонах можно изменять параметры системы, не нарушая ее устойчивости. Вторая постановка задачи об устойчивости имеет место при наладке и эксплуатации систем автоматического управления.

Анализ систем на устойчивость и выбор средств для стабилизации неустойчивых систем осуществляются с использованием критериев (признаков) устойчивости, которые формулируются исходя из общих условий устойчивости.

Вопрос об устойчивости системы сводится к выяснению устойчивости свободного движения системы и требует анализа характера решения уравнения свободного движения, составленного относительно отклонения выходной величины x от установившегося состояния. На основании (2.35) и (2.37) уравнение свободного движения системы может быть приведено к виду

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_{n-1} \frac{dx}{dt} + a_n x = 0, \quad (2.61)$$

где n — порядок уравнения системы; a_i — коэффициенты, определяемые из параметров системы.

Линейное однородное дифференциальное уравнение (2.61) имеет общее решение вида

$$x = c_1 e^{p_1 t} + c_2 e^{p_2 t} + \dots + c_n e^{p_n t}, \quad (2.62)$$

где c_1, c_2, c_n — постоянные интегрирования; p_1, p_2, p_n — корни характеристического уравнения.

Характеристическое уравнение, соответствующее дифференциальному уравнению (2.61), имеет вид $K(p) + H(p) = 0$, где $K(p)$, $H(p)$ — характеристические полиномы соответственно числителя и знаменателя передаточной функции разомкнутой системы.

Если анализируется устойчивость разомкнутой системы, то, как следует из (2.36), характеристическое уравнение имеет вид $H(p) = 0$. Чаще приходится анализировать устойчивость замкнутой системы.

Во всех случаях характеристическое уравнение свободного движения приводится к виду

$$a_0 p^n + a_1 p^{n-1} + \dots + a_i p^{n-i} + \dots + a_{n-1} p + a_n = 0. \quad (2.63)$$

Корни уравнения (2.63) могут быть либо вещественными, либо комплексными попарно сопряженными.

В случае вещественных корней обозначим $p_i = \lambda_i$. Тогда общее решение уравнения (2.61) представляется как сумма n решений вида $x_i = c_i e^{\lambda_i t}$. Так как в (2.61) величина x означает отклонение регулируемой величины от установившегося состояния, то очевидное условие устойчивости при $t \rightarrow \infty$ имеет вид $x(t) \rightarrow 0$. Каждая составляющая общего решения стремится к нулю лишь при отрицательном значении вещественного корня, т. е. при $\lambda_i < 0$.

Комплексные корни имеют вид

$$p_k = \alpha_k \pm j\omega_k.$$

где α_k — вещественная часть комплексного корня; ω — коэффициент при мнимой части — угловая частота колебания.

Каждому корню p_k в рассматриваемом случае будет соответствовать частное решение

$$x_k = X_k \sin(\omega_k t + \varphi_k), \quad (2.64)$$

где $X_k = B_k e^{\alpha_k t}$ — амплитуда составляющей x_k ; B_k , φ_k — постоянные интегрирования.

Движение, описываемое уравнением (2.64), будет сходящимся, если при $t \rightarrow \infty$ решение $x_k(t) \rightarrow 0$. Это возможно лишь при отрицательном значении вещественной части комплексного корня.

Исходя из изложенного, можно вывести условие устойчивости линейных динамических систем: линейная система будет устойчива, если все вещественные корни и все вещественные части комплексных корней характеристического уравнения, соответствующего исходному дифференциальному уравнению свободного движения системы, будут отрицательными.

Весьма важной для практического анализа устойчивости является геометрическая интерпретация условий устойчивости. Она сводится к тому, что в устойчивой линейной системе все корни характеристического уравнения, соответствующего дифференциальному уравнению свободного движения системы, на комплексной плоскости в координатах $\alpha - j\omega$ должны располагаться слева от мнимой оси. Такие корни впредь будем называть «левыми». Следовательно, условие устойчивости в такой постановке сводится к тому, чтобы корни характеристического уравнения были только левыми. Исходя из этого, в теории управления разработан ряд признаков — критериев устойчивости линейных динамических систем, которые позволяют вынести суждения об устойчивости системы, не вычисляя значения корней характеристического уравнения. В практике исследования линейных систем автоматического управления широко используются две группы критериев устойчивости — алгебраические и частотные. Применение того или иного критерия определяется условием конкретной задачи, например порядком дифференциального уравнения свободного движения системы, наличием или отсутствием тех или иных характеристик исследуемой системы — передаточных функций, АФХ, ЛАФЧХ и др.

Не останавливаясь на доказательствах, которые построены на утверждении необходимости левых корней для обеспечения устойчивости линейной динамической системы, приведем некоторые из критериев устойчивости.

Критерий Рауса. Практическое использование алгебраического критерия Рауса сводится к тому, что на базе значений коэффициентов характеристического уравнения (2.63) составляется таблица Рауса (табл. 2.3), коэффициенты которой b_{re} назовем элементами таблицы Рауса, где r — номер строки таблицы; e — номер столбца.

Элементы первой строки таблицы являются коэффициентами характеристического уравнения (2.63) с четными индексами, т. е. $b_{11} = a_0$; $b_{12} = a_2$ и т. д. Элементы второй строки представляют собой коэффициенты уравнения (2.63) с нечетными индексами, т. е. $b_{21} =$

Таблица 2.3

b_{11}	b_{12}	b_{13}	b_{14}	...	...
b_{21}	b_{22}	b_{23}	b_{24}	...	...
b_{31}	b_{32}	b_{33}	b_{34}	...	...
...	...	...	...	...	...
...	...	...	...	...	...

$=a_1$; $b_{22}=a_3$ и т. д. Для нахождения остальных элементов таблицы Рауса можно воспользоваться выражением

$$b_{re} = \frac{b_{r-1, 1} b_{r-2, e+1} - b_{r-2, 1} b_{r-1, e+1}}{b_{r-1, 1}}.$$

Критерий Рауса формулируется так: линейная динамическая система будет устойчивой, если все элементы первого столбца таблицы Рауса положительны. Форма критерия Рауса позволяет легко реализовать его с помощью цифровых ЭВМ. Однако этот критерий неудобен для анализа влияния отдельных параметров на устойчивость системы.

Критерий Гурвица. Для анализа устойчивости линейной системы по критерию Гурвица из коэффициентов характеристического уравнения (2.63) составляются определители Гурвица:

$$\Delta_1 = a_1; \quad \Delta_2 = \begin{vmatrix} a_1 a_0 \\ a_3 a_2 \end{vmatrix}; \quad \Delta_3 = \begin{vmatrix} a_1 a_0 & 0 \\ a_3 a_2 & a_1 \\ a_5 a_4 & a_3 \end{vmatrix}; \quad \Delta_4 = \begin{vmatrix} a_1 a_0 & 0 & 0 \\ a_3 a_2 & a_1 & 0 \\ a_5 a_4 & a_3 & a_2 \\ a_7 a_6 & a_5 & a_4 \end{vmatrix}.$$

Число определителей Гурвица равно порядку характеристического уравнения n .

По критерию Гурвица линейная динамическая система будет устойчивой, если все коэффициенты характеристического уравнения и все определители Гурвица положительны.

Любой определитель Δ_i составляется по одному общему правилу: число строк и столбцов определителя Δ_i равно i ; по диагонали определителя располагаются подряд коэффициенты характеристического уравнения от a_1 до a_i ; влево от диагонали на каждой строке располагаются коэффициенты с возрастающими индексами, вправо — с убывающими. Элементы, располагающиеся вправо от a_0 , и элементы с индексами, превышающими степень характеристического уравнения, заменяются нулями.

Из критерия Гурвица следуют простые алгебраические выражения, составленные из коэффициентов характеристического уравнения, удобные для анализа устойчивости систем невысокого порядка. Так, для систем первого и второго порядков достаточным условием устойчивости является положительность всех коэффициентов характеристического уравнения. Для систем более высокого порядка это условие сохраняется, но, кроме того, необходимо выполнение следующих соотношений:

в системах третьего порядка $a_1 a_2 - a_0 a_3 > 0$;
 в системах четвертого порядка $(a_1 a_2 - a_0 a_3) a_3 - a_1^2 a_4 > 0$.

Недостаток критерия Гурвица — громоздкость вычислений определителей для систем высокого порядка.

Частотный критерий устойчивости — критерий Михайлова

Частотные критерии базируются на свойствах частотных характеристик устойчивых систем. Большую роль в развитии теории устойчивости сыграл частотный критерий устойчивости, предложенный в 1936 г. А. В. Михайловым. Так же, как и алгебраические критерии, частотные критерии вытекают из неперемennого условия наличия только левых корней в характеристическом уравнении устойчивой линейной динамической системы. Не останавливаясь на доказательстве критерия А. В. Михайлова, рассмотрим его практическое использование для анализа устойчивости. С этой целью характеристическое уравнение запишем в виде

$$a_0 p^n + a_1 p^{n-1} + \dots + a_{n-1} p + a_n = F(p). \quad (2.65)$$

Положив в (2.65) $p = j\omega$ и отделяя вещественную часть от мнимой, полином $F(p)$ приведем к виду

$$F(j\omega) = U(\omega) + jV(\omega), \quad (2.66)$$

где $U(\omega)$ — вещественная часть — сумма всех членов, содержащих j в четных степенях; V — мнимая часть выражения (2.66).

В соответствии с критерием Михайлова условие устойчивости

$$\Delta \arg F(j\omega) = n \frac{\pi}{2}, \quad 0 < \omega < \infty. \quad (2.67)$$

Геометрическое место точек конца вектора $F(j\omega)$ при изменении частоты в диапазоне $0 < \omega < \infty$ называется *годографом вектора*, или *годографом Михайлова*. В соответствии с (2.67) критерий Михайлова формулируется следующим образом: динамическая система, описываемая линейным дифференциальным уравнением n -го порядка, устойчива, если при изменении ω от 0 до ∞ годограф вектора $F(j\omega)$ последовательно проходит в положительном направлении (против часовой стрелки) n квадрантов комплексной плоскости. На рис. 2.6 приведены примеры годографов устойчивых и неустойчивых систем.

Амплитудно-фазовый критерий устойчивости — критерий Найквиста. Этот критерий основан на связи свойства устойчивости замкнутой системы с формой АФХ разомкнутой устойчивой системы. Вопрос же об устойчивости разомкнутой системы решается обычно легко. В соответствии с критерием Михайлова разомкнутая устойчивая система с передаточной функцией $W_p(p) = K_p(p)/H_p(p)$ отвечает условию

$$\Delta \arg H_p(j\omega) = n \frac{\pi}{2}, \quad 0 \leq \omega \leq \infty, \quad (2.68)$$

где n — порядок характеристического полинома $H_p(p) = 0$.

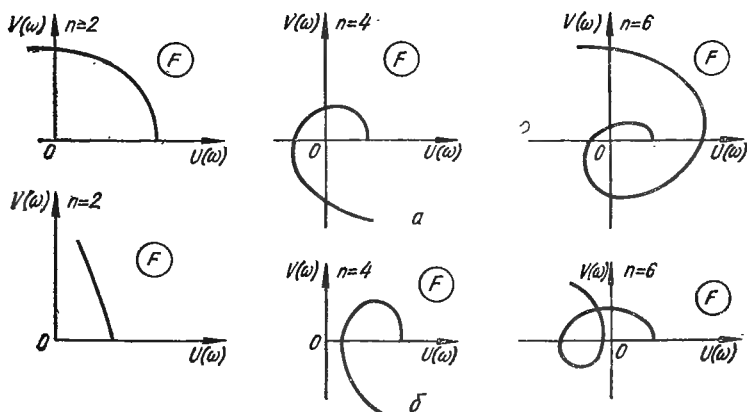


Рис. 2.6. Годографы систем:
а — устойчивых; б — неустойчивых

АФХ замкнутой системы описывается выражением

$$W_s(j\omega) = W_p(j\omega) / [1 + W_p(j\omega)].$$

Обозначим знаменатель полученного выражения через $W_1(j\omega)$, т. е.

$$W_1(j\omega) = 1 + W_p(j\omega) = 1 + \frac{K_p(j\omega)}{H_p(j\omega)} = \frac{H(j\omega)}{H_p(j\omega)}, \quad (2.69)$$

где $H(j\omega) = K_p(j\omega) + H_p(j\omega)$ — характеристический полином замкнутой системы при $p = j\omega$.

В соответствии со свойствами передаточных функций порядок полинома $H(p)$ не превышает порядка полинома $H_p(p)$, так как $H(p) = K_p(p) + H_p(p)$, а порядок полинома $K_p(p)$ меньше порядка полинома $H_p(p)$.

Поэтому критерий Михайлова для замкнутой системы можно записать в виде

$$\Delta \arg H(j\omega) = (n - 2m) \frac{\pi}{2}, \quad 0 \leq \omega \leq \infty, \quad (2.70)$$

где m — число правых корней системы, имеющей в замкнутом состоянии характеристический полином $H(p) = 0$.

Коэффициент 2 перед m введен потому, что каждый правый корень не только не обеспечивает положительного поворота вектора, но и создает его отрицательный поворот (по часовой стрелке).

Из (2.69) вытекает, что

$$\Delta \arg W_1(j\omega) = \Delta \arg H(j\omega) - \Delta \arg H_p(j\omega).$$

Или, учтя (2.68) и (2.70), получаем

$$\Delta \arg W_1(j\omega) = (n - 2m) \frac{\pi}{2} - n \frac{\pi}{2} = -m\pi.$$

В устойчивой замкнутой системе правых корней в характерис-

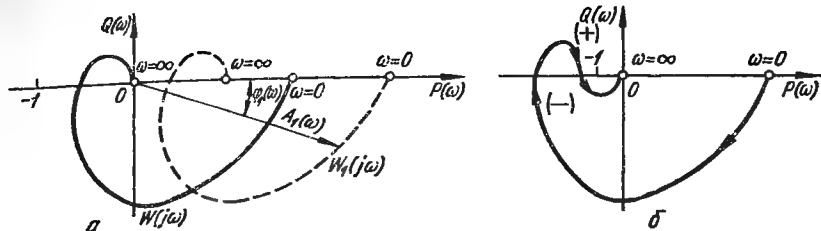


Рис. 2.7. Амплитудно-фазовый критерий устойчивости (критерий Найквиста):
а — частный случай; б — общий

тическом уравнении нет, т. е. $m=0$, следовательно, условием устойчивости замкнутой системы будет

$$\Delta \arg W_1(j\omega) = 0. \quad (2.71)$$

Условие (2.71) выполняется только тогда, когда кривая $W_1(j\omega)$ при изменении частоты от 0 до ∞ не охватывает начала координат в комплексной плоскости $P(\omega)$; $jQ(\omega)$, где $P(\omega)$ — вещественная часть выражения $W_1(j\omega)$, $Q(\omega)$ — мнимая часть (рис. 2.7). Действительно, только в этом случае результирующий поворот вектора $W_1(j\omega)$ при изменении ω от 0 до ∞ будет равен нулю, так как возрастание угла $\varphi(\omega)$, обусловленное движением вектора $W_1(j\omega)$ в положительном направлении (против часовой стрелки), будет компенсироваться таким же убыванием $\varphi(\omega)$, обусловленным движением вектора $W_1(j\omega)$ в отрицательном направлении (по часовой стрелке).

Как видно из (2.69), переход на комплексной плоскости от годографа вектора $W_1(j\omega)$ к годографу вектора АФХ разомкнутой системы $W_p(j\omega)$ осуществляется сдвигом кривой $W_1(j\omega)$ влево на -1 , так как $W_p(j\omega) = W_1(j\omega) - 1$.

Выполнив эту операцию, получаем следующую формулировку амплитудно-фазового критерия устойчивости — критерия Найквиста: линейная динамическая система, устойчивая в разомкнутом состоянии, устойчива и в замкнутом состоянии, если АФХ разомкнутой системы при изменении частоты от 0 до ∞ не охватывает на комплексной плоскости точку с координатами $(-1; j0)$ (рис. 2.7, а).

Более общая формулировка критерия Найквиста относится к системам, имеющим так называемую АФХ второго рода (рис. 2.7, б), когда $W_p(j\omega)$ несколько раз пересекает вещественную ось левее точки $P(\omega) = -1$. В этом случае условимся считать положительным переход годографа через вещественную ось, если он совершается сверху вниз, и отрицательным, если он совершается снизу вверх. Тогда формулировка критерия Найквиста принимает вид: линейная динамическая система, устойчивая в разомкнутом состоянии, устойчива и в замкнутом состоянии, если при изменении частоты от 0 до $+\infty$ разность между числом положительных переходов годографа АФХ разомкнутой системы через участок вещественной оси $(-1, -\infty)$ и числом отрицательных переходов равна нулю. Из этого условия видно, что система, устойчивая в разомк-

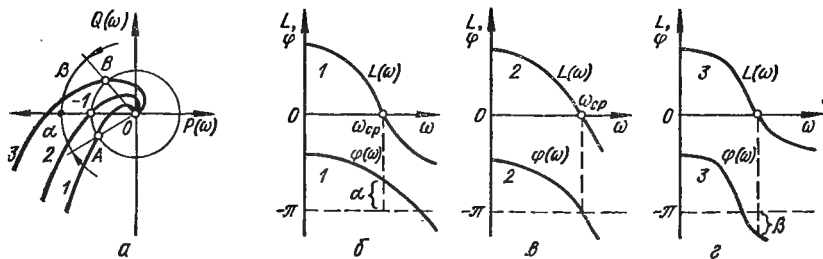


Рис. 2.8. Логарифмический критерий устойчивости

нудом состоянии и имеющая АФХ в форме кривой, показанной на рис. 2.7, б, устойчива и в замкнутом состоянии.

Логарифмический критерий устойчивости. Этот критерий основан на связи свойств устойчивой замкнутой динамической системы с формой ЛАФЧХ разомкнутой системы (рис. 2.8).

В соответствии с критерием Найквиста устойчивая разомкнутая система, имеющая АФХ в форме кривой 1, устойчива с запасом по фазе, равным α . При этом под запасом по фазе понимается угол между отрицательным направлением вещественной оси и прямой, соединяющей начало координат с точкой А — точкой пересечения АФХ разомкнутой системы с окружностью единичного радиуса.

Система, имеющая в разомкнутом состоянии АФХ в форме кривой 2, будучи устойчивой в разомкнутом состоянии, в замкнутом состоянии находится на границе устойчивости (запас по фазе равен нулю). Наконец, разомкнутая устойчивая система, имеющая АФХ в форме кривой 3, в замкнутом состоянии будет неустойчивой с недостатком фазы, равном β . Нетрудно заметить, что этим трем случаям соответствуют логарифмические амплитудно-фазовые частотные характеристики, помеченные теми же индексами, что и АФХ. Анализируя формы ЛАФЧХ для устойчивых и неустойчивых систем, можно сформулировать следующий логарифмический критерий устойчивости: динамическая линейная система, устойчивая в разомкнутом состоянии, устойчива и в замкнутом состоянии, если на всем диапазоне частот, на котором $L(\omega) > 0$, значение фазы $\varphi(\omega) > -180^\circ$ или (равносильно) на всем диапазоне частот $0 < \omega < \omega_{ср}$ фазовая характеристика не пересекает линии $-\pi$. Под частотой среза $\omega_{ср}$ понимается частота, при которой $L(\omega)$ пересекает ось частот.

Для устойчивых разомкнутых линейных систем, имеющих АФХ второго рода (рис. 2.7, б), логарифмический критерий устойчивости может быть сформулирован так: линейная динамическая система, устойчивая в разомкнутом состоянии, будет устойчивой и в замкнутом, если на всем диапазоне частот, на котором $L(\omega) > 0$, количество положительных переходов фазовой характеристики $\varphi(\omega)$ через линию $-\pi$ равно количеству отрицательных переходов (рис. 2.9).

Выделение областей устойчивости. Критерий Вышнеградского.

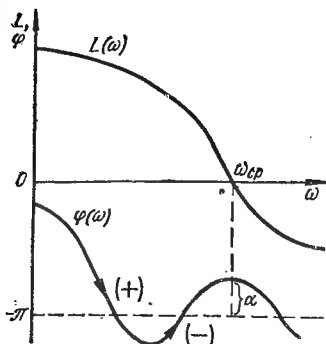


Рис. 2.9. Общий случай логарифмического критерия устойчивости

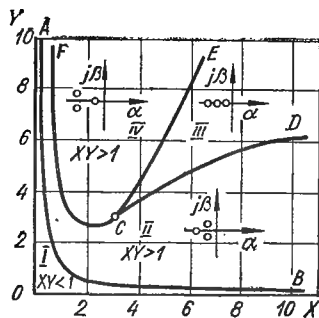


Рис. 2.10. Диаграмма Вышнеградского

При наладке и эксплуатации линейных динамических систем возникает необходимость изменять отдельные параметры системы, следовательно, неоднократно проверять систему на устойчивость. Поэтому важное значение имеют методы, позволяющие определить область изменения тех или иных параметров, допустимую с точки зрения условий устойчивости. Так как коэффициенты характеристического уравнения системы определяются через ее параметры, то поставленная задача может рассматриваться как задача определения допустимого по условиям устойчивости диапазона изменения коэффициентов характеристического уравнения.

Такая задача для диапазона изменения двух параметров была впервые поставлена и решена русским инженером и ученым И. А. Вышнеградским. Результат решения известен под названием *диаграммы Вышнеградского*, широко и успешно применяемой в практике анализа и синтеза линейных систем автоматического управления.

Критерий Вышнеградского обоснован для систем, описываемых линейными уравнениями третьего порядка, но идеи, заложенные в этом критерии, оказали большое влияние на развитие теории устойчивости линейных динамических систем.

Пусть имеется система, характеристическое уравнение которой имеет вид

$$a_0 p^3 + a_1 p^2 + a_2 p + a_3 = 0 \quad (2.72)$$

или

$$p^3 + c_1 p^2 + c_2 p + c_3 = 0, \quad (2.73)$$

где $c_1 = a_1/a_0$; $c_2 = a_2/a_0$; $c_3 = a_3/a_0$.

Сделаем подстановки: $p = \sqrt[3]{Z/c_3}$; $c_1 = X\sqrt[3]{c_3}$; $c_2 = Y\sqrt[3]{c_3^2}$

или $Z = p/\sqrt[3]{c_3}$; $X = c_1/\sqrt[3]{c_3}$; $Y = c_2/\sqrt[3]{c_3^2}$.

С учетом этих обозначений уравнение (2.73) примет вид

$$Z^3 + XZ^3 + YZ + 1 = 0.$$

Если система находится на границе устойчивости, то уравнение (2.73) будет иметь один действительный отрицательный корень и два мнимых корня:

$$p_1 = -\alpha; \quad p_2 = -j\beta; \quad p_3 = j\beta.$$

Левую часть уравнения (2.73) можно разложить на множители:

$$\begin{aligned} p^3 + c_1 p^2 + c_2 p + c_3 &= (p + \alpha)(p - j\beta)(p + j\beta) = \\ &= (p + \alpha)(p^2 + \beta^2) = p^3 + p^2\alpha + p\beta^2 + \alpha\beta^2. \end{aligned}$$

Приравнивая в этом уравнении коэффициенты при одинаковых степенях p , находим $c_1 = \alpha, c_2 = \beta^2, c_3 = \alpha\beta^2$.

С учетом этих соотношений справедливо

$$c_1 c_2 - c_3 = 0. \quad (2.74)$$

Выразив в (2.74) коэффициенты через X и Y , получим

$$XY - 1 = 0, \text{ т. е. } XY = 1. \quad (2.75)$$

Это уравнение является уравнением гиперболы, коэффициенты которого выражены через коэффициенты характеристического уравнения системы, т. е. через ее параметры. При этом, как следует из приведенных рассуждений, все точки в плоскости Y, Z , лежащие на гиперболе (2.75), соответствуют таким значениям параметров системы, при которых эта система находится на границе устойчивости. Геометрическое место точек в плоскости параметров линейной динамической системы, описываемой дифференциальным уравнением третьего порядка, соответствующее состоянию системы на границе устойчивости, называется *гиперболой Вышнеградского* — уравнение (2.75) и рис. 2.10. Гипербола Вышнеградского разбивает плоскость Y, X на две области: $XY < 1$ (область I) и $XY > 1$ (область II). Для выяснения вопроса об устойчивости системы с параметрами, лежащими в областях I и II, примем в уравнении (2.73) $c_3 = 0$. Тогда получим: $p = 0; p^2 + c_1 p + c_2 = 0$. Второе уравнение будет иметь корни с отрицательными вещественными частями при условии $c_1 > 0; c_2 > 0$. Учитывая, что $c_3 = 0$, приходим к выводу, что два корня имеют отрицательные вещественные части при условии

$$c_1 c_2 - c_3 > 0. \quad (2.76)$$

Это условие остается справедливым и в случае, когда c_3 отличается от нуля, так как при малом изменении c_3 корни уравнения (2.73) изменяются мало, согласно теореме о непрерывной зависимости корней характеристического уравнения от его коэффициентов. Следовательно, условие (2.76) является условием устойчивости системы, которое, с учетом значений c_1, c_2 и c_3 , может быть заменено условием $XY > 1$.

Таким образом, область параметров системы, расположенная выше гиперболы Вышнеградского (рис. 2.10), соответствует области устойчивости, а область, расположенная ниже гиперболы, — неустойчивая область.

На основании сказанного критерий устойчивости Вышнеградского формулируется так: линейная динамическая система, описываемая дифференциальным уравнением третьего порядка, устой-

чива, если при положительных коэффициентах характеристического уравнения выполняется условие $XY > 1$.

Устойчивую область на диаграмме Вышнеградского можно разбить на ряд подобластей с одинаковым характером устойчивого переходного процесса. Например, анализ уравнения (2.73) из условий наличия трех вещественных отрицательных корней позволяет выделить кривую ECD , описываемую уравнением $4(x^3 + y^3) - x^2y^2 - 18xy + 27 = 0$, которая на диаграмме выделяет область III. В пределах этой области параметры системы обеспечивают отрицательные вещественные корни, т. е. обеспечивают аperiodический переходный процесс в системе.

Рассуждая аналогично, можно получить уравнение кривой FCE ($2x^3 - 9xy + 27 = 0$), выделяющей в плоскости X, Y область IV с параметрами системы, обеспечивающими один вещественный отрицательный корень и два комплексных корня с отрицательной вещественной частью. При этом вещественные части комплексных корней по абсолютному значению больше вещественного корня, что соответствует монотонным переходным процессам. Наконец, в области II имеют место один вещественный отрицательный корень и два комплексных корня с отрицательной вещественной частью. При этом вещественный корень по абсолютному значению больше вещественной части комплексных корней, что соответствует колебательным переходным процессам.

Таким образом, диаграмма Вышнеградского представляет собой плоскость двух параметров линейной динамической системы третьего порядка с выделенными устойчивыми и неустойчивыми областями параметров системы и областями, соответствующими аperiodическим и колебательным переходным процессам. По диаграмме легко определяется диапазон изменения двух параметров линейной динамической системы, обеспечивающий тот или иной переходный процесс в системе.

Выделение областей устойчивости в плоскости параметров линейной динамической системы, описываемой дифференциальным уравнением n -го порядка, осуществляется на базе общего метода D -разбиения, рассматриваемого в специальных курсах по теории автоматического управления. Этот метод базируется на анализе характеристического уравнения линейной системы. Он позволяет сравнительно легко определить допустимые с точки зрения устойчивости диапазоны изменения одного или двух коэффициентов характеристического уравнения, выражаемых через параметры системы.

2.5. КАЧЕСТВО УПРАВЛЕНИЯ В ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМАХ

Под *качеством управления* понимается характер переходных процессов в системе, вызванных управляющими и возмущающими воздействиями. Законы изменения величин, характеризующих систему, задаются. Исследование качества процесса управления сводится, таким образом, к определению действительного характера

переходных процессов в системе и сопоставлению его с заданными законами. Если система не удовлетворяет требуемым показателям качества управления, то в нее вводятся дополнительные элементы — корректирующие устройства, которые, не нарушая основного назначения системы, обеспечивают требуемое качество управления.

Для оценки качества управления необходимо располагать количественными или качественными показателями, характеризующими формы переходных процессов в системе.

Оценки качества управления могут быть косвенными или прямыми.

Косвенные оценки получают на основе анализа тех или иных характеристик системы, например частотных. Косвенные методы оценки позволяют получить некоторые параметры переходного процесса, например установить его принципиальный характер — сходящийся он или расходящийся.

Среди косвенных методов оценки качества процесса управления получили распространение корневые, интегральные и частотные методы.

Корневые методы вытекают из наличия связи между формой переходного процесса в системе и характером распределения на комплексной плоскости корней характеристического уравнения. Например, ранее было показано, что если имеется хоть один правый корень, то переходный процесс в системе будет расходящимся. Корневые методы наиболее эффективны, если удастся выполнить анализ не только полюсов, но и нулей передаточной функции исследуемой системы. При этом под полюсами понимаются корни знаменателя передаточной функции, т. е. характеристического полинома $H(p)=0$, а под нулями — корни числителя передаточной функции, т. е. полинома $K(p)=0$.

Вместе с тем корневые методы наиболее полно разработаны для систем с передаточными функциями вида $W(p)=k/H(p)$, т. е. с передаточными функциями, не содержащими нулей.

В *интегральных методах* оценками качества служат определенные интегралы по времени от некоторой функции физических величин, характеризующих системы, вычисляемые без решения дифференциального уравнения системы. Интегральные методы наиболее удобны для сравнительной оценки обобщенных показателей качества в системах управления, близких по некоторым свойствам.

Частотные методы базируются на связи между формой переходного процесса в системе и вещественной частью частотной характеристики. Эти методы позволяют приближенно построить всю кривую переходного процесса, не прибегая к решению дифференциальных уравнений, описывающих исследуемую систему.

К *прямым оценкам* относятся количественные показатели формы переходного процесса физических переменных в системе при заданных воздействиях на нее. Прямая оценка качества может быть получена лишь после того, как построена интересующая нас зависимость $x(t)$.

Построение переходных процессов операторным методом. Операторный метод анализа качества управления базируется на ис-

пользовании передаточной функции системы, из которой можно получить изображение искомой величины

$$x(p) = K(p)/H(p)F(p), \quad (2.77)$$

где $F(p)$ — преобразованная по Лапласу функция воздействия на систему (возмущающего или управляющего).

Располагая изображением искомой переменной $x(p)$ и применяя обратное преобразование Лапласа, можно определить переходный процесс в системе, т. е. функцию $x(t)$.

Выражение (2.77) позволяет найти изображение функции переходного процесса лишь при нулевых начальных условиях, что вытекает из определения передаточной функции системы. Как известно, начальные условия будут нулевыми тогда, когда в начальный момент воздействия на систему сама функция $x(t)$ и все ее производные до $(n-1)$ -й включительно равны нулю. Начальные условия зависят от состояния системы в момент приложения возмущающего воздействия и от его формы, которая определяется видом полинома $K(p)$ — полинома числителя передаточной функции системы. Можно выделить следующие обстоятельства, обеспечивающие нулевые начальные условия: 1) начальные условия будут нулевыми при любом воздействии на систему, если в момент его приложения система находилась в равновесном состоянии, а полином $K(p)=k$, т. е. не содержит p ; 2) система в момент приложения воздействия находилась в равновесном состоянии, в числе членов полинома $K(p)$ имеются такие, которые содержат p , но все производные функции воздействия на систему, вплоть до m -й включительно, при $t=0$ равны нулю (здесь m — порядок полинома $K(p)$); 3) система находилась в неравновесном состоянии, полином $K(p)$ содержит члены, содержащие p , но форма возмущающего воздействия такова, что при $t=0$ каждый член правой части дифференциального уравнения системы, содержащий производную воздействия на систему до m -й включительно, уравнивается соответствующим членом левой части уравнения, содержащим производную возмущения того же порядка. При ненулевых начальных условиях изображение выходной величины системы имеет вид

$$x(p) = \frac{K(p)}{H(p)} F(p) + \frac{M(p)}{H(p)}, \quad (2.78)$$

где $M(p)$ — полином, учитывающий влияние ненулевых начальных условий.

Для множества случаев изображение функции воздействия на систему имеет вид дробно-рационального выражения $F(p) = F_1(p)/F_2(p)$. Тогда (2.78) принимает вид

$$x(p) = \frac{K(p)F_1(p) + F_2(p)M(p)}{H(p)F_2(p)}.$$

Линейные динамические системы часто исследуются в режиме воздействия в форме единичного скачка $\bar{f}$:

$$\bar{f} = f(t) \begin{vmatrix} 0 & \text{при } t < 0 \\ F & \text{при } t > 0 \end{vmatrix}.$$

В этом случае

$$F(p) = \bar{f}/p; \quad F_1(p) = \bar{f}; \quad F_2(p) = p.$$

Тогда (2.78) примет вид

$$x(p) = \frac{K(p)\bar{f} + pM(p)}{pH(p)}. \quad (2.79)$$

Если в (2.79) корни полинома $H(p) = 0$ будут простыми (не кратными) и при $t=0$ система находится в равновесном состоянии, то, используя формулу разложения Хэвисайда, получим

$$x(t) = \frac{K(0)}{H(0)} + \sum_{k=1}^n \left[\frac{K(p)}{pH'(p)} \right]_{p=p_k} e^{p_k t}, \quad (2.80)$$

где

$$H'(p) = \left[\frac{d}{dt} H(p) \right]_{p=p_k};$$

n — порядок полинома $H(p) = 0$; $p_1, p_2, \dots, p_n$ — корни полинома $H(p) = 0$; $K(0) = K(p)$ и $H(0) = H(p)$ при $p = 0$.

На основании свойств полинома $H(p) = 0$, его можно представить в виде $H(p) = a_0(p - p_1) \dots (p - p_k) \dots (p - p_n) = a_0 \prod_{k=1}^n (p - p_k)$; $H'(p) = (p_k - p_1) \dots (p_k - p_n) = \prod_{q=1}^{q=n-1} (p_k - p_q)$.

Тогда (2.80) можно переписать так:

$$x(t) = \frac{K(0)}{H(0)} + \sum_{k=1}^n \frac{K(p_k)}{p_k \prod_{q=1, q \neq k}^n (p_k - p_q)} e^{p_k t}. \quad (2.81)$$

Выражение (2.81) позволяет построить переходный процесс $x(t)$ в линейной динамической системе при единичном воздействии.

Операторный метод построения переходных процессов становится неудобным для систем высокого порядка, так как требуется большой объем вычислений.

Построение переходных процессов частотным методом. Метод базируется на использовании частотной характеристики $P(\omega)$ — вещественной части выражения передаточной функции $W(p)$ системы при $p = j\omega$:

$$W(j\omega) = P(\omega) + jQ(\omega).$$

Не останавливаясь на доказательстве, укажем, что между функциями $x(t)$ и $P(\omega)$ имеют место следующие зависимости:

$$x(t) = \frac{2}{\pi} \int_0^{\infty} \frac{P(\omega)}{\omega} \sin \omega t d\omega \quad (2.82)$$

или

$$x(t) = P(0) + \frac{2}{\pi} \int_0^{\infty} \frac{Q(\omega)}{\omega} \cos \omega t d\omega. \quad (2.83)$$

Функции $P(\omega)$ и $Q(\omega)$ имеют сложный характер, вследствие

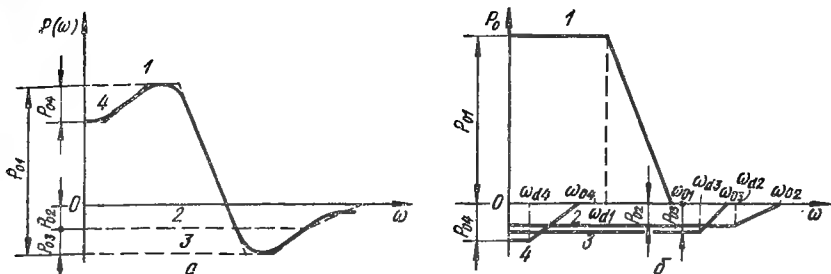


Рис. 2.11. Вещественная частотная характеристика и ее составляющие

чего их трудно описать аналитически. Поэтому переходные процессы непосредственно по выражениям (2.82) и (2.83) на практике не рассчитываются. Однако расчет существенно упрощается, если кривую $P(\omega)$ (рис. 2.11) заместить совокупностью трапеций:

$$P(\omega) = \sum_{i=1}^n P_i(\omega),$$

где i — номер трапеции.

Каждой составляющей $P_i(\omega)$ соответствует переходный процесс вида

$$x_i(t) = \frac{2}{\pi} \int_0^{\infty} \frac{P_i(\omega)}{\omega} \sin \omega t d\omega.$$

Таким образом,

$$x(t) = \sum_{i=1}^n x_i(t). \quad (2.84)$$

Выражение (2.84) лежит в основе метода расчета переходных процессов в линейных динамических системах с использованием вещественной частотной характеристики $P(\omega)$. Для практической реализации метода каждую i -ю трапецию (рис. 2.11) определим следующими параметрами: P_{oi} — высота; ω_{oi} — интервал пропуска частот; ω_{di} — интервал равномерного пропуска частот; $v = \omega_{di}/\omega_{oi}$ — коэффициент наклона. Кроме того, вводится понятие типовой единичной трапеции, для которой принято $P_0 = 1$, $\omega_0 = 1$ при любом значении v .

Зависимость (2.82) для случая, когда $P(\omega)$ замещается типовой единичной трапецией, дает следующее выражение переходного процесса:

$$h(\tau) = \frac{2}{\pi(1-v)} \left[\text{Si}(\tau) - v \text{Si}(v\tau) + \frac{\cos \tau - \cos v\tau}{\tau} \right], \quad (2.85)$$

где $\text{Si}(\tau)$ — интегральной синус выражения $\int_0^{\infty} \frac{\sin \tau}{\tau} d\tau$.

Функция $h_i(\tau)$ легко преобразуется в соответствующую функцию $x_i(t)$ путем пересчета масштабов по осям. На основании связи

характера функции $P(\omega)$ и формы переходного процесса этот пересчет осуществляется так:

$$x_i(t) = h_i(\tau) P_{0i}; \quad t = \tau/\omega_{0i}. \quad (2.86)$$

Значения функции $h(\tau)$ для различных значений ν и τ рассчитаны и сведены в таблицы, которые позволяют легко получить точные кривой $x_i(t)$.

На основании изложенного порядок построения графика переходного процесса в линейной динамической системе при единичном воздействии на входе с использованием вещественной частотной характеристики сводится к такой последовательности операций:

1) на основе передаточной функции системы $W(p)$ путем замены $p=j\omega$ и отделения вещественной части от мнимой находится характеристика $P(\omega)$;

2) функция $P(\omega)$ разбивается на трапециевидные составляющие, для каждой из которых определяются параметры P_{0i} , ω_{0i} , ω_{di} и ν ;

3) с помощью таблиц находятся h -функции для относительного времени τ ;

4) используя соотношения масштабов, определенные выражениями (2.86), осуществляется переход от функции $h_i(\tau)$ к функции $x_i(t)$;

5) определяется кривая переходного процесса в системе как алгебраическая сумма кривых $x_i(t)$:

$$x(t) = \sum_{i=1}^n x_i(t),$$

где n — число трапеций, которыми замещена характеристика $P(\omega)$.

Рассмотренный метод можно распространить на построение переходных процессов при воздействии, отличающемся от единичного, о чем пойдет речь ниже.

Как, отмечалось, переходные процессы в линейных динамических системах строятся обычно для замкнутых систем. Поэтому наиболее трудоемкой частью расчета является определение характеристики $P(\omega)$ замкнутой системы. Эта часть расчета существенно упрощается при использовании для определения вещественной частотной характеристики $P(\omega)$ замкнутой системы характеристик $L(\omega)$ и $\nu(\omega)$ разомкнутой системы. В теории автоматического управления широко используются номограммы для определения функции $P(\omega)$ замкнутой системы по логарифмическим амплитудно-фазовым частотным характеристикам разомкнутой системы.

От переходного процесса, вызванного единичным входным воздействием, можно перейти к переходному процессу от произвольного воздействия $u(t)$, используя интеграл Дюамеля, или свертки:

$$x(t) = \int_0^t u(t-\tau) \omega(\tau) d\tau \quad (2.87)$$

или

$$x(t) = \int_0^t \omega(t-\tau) u(\tau) d\tau,$$

где τ — момент приложения воздействия $u(t)$; $\omega(t)$ — импульсная переходная функция.

Импульсная переходная функция $\omega(t)$ представляет собой переходный процесс в системе, к которой приложено импульсное воздействие в виде δ -функции, отвечающей условию

$$\int_0^{\infty} \delta(t) dt = 1; \quad \delta(p) = 1, \quad (2.88)$$

при этом $\delta(t)$ существует только при $t=0$.

Нетрудно убедиться, что между импульсной переходной функцией $\omega(t)$ и функцией $h(t)$ существует зависимость

$$\omega(t) = \dot{h}(t), \quad (2.89)$$

т. е. $\omega(t)$ представляет собой первую производную по времени h -функции — функции переходного процесса при единичном возмущении.

2.6. КОРРЕКЦИЯ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

Линейную динамическую систему, элементы которой рассчитаны из условий обеспечения статических режимов работы, назовем *исходной системой*. Исходная система может оказаться неудовлетворительной с точки зрения требований динамики, т. е. неустойчивой или устойчивой, но не обеспечивающей требуемых форм переходных процессов. В таких случаях исходная система дополняется специальными элементами — корректирующими устройствами (КУ), которые вводятся в систему с целью придания ей необходимых динамических свойств, не изменяя основного назначения системы, определяемого условиями работы объекта управления.

По характеру включения корректирующих устройств в исходную систему их можно разделить на две группы — последовательные корректирующие устройства (КУ) и параллельные. Первые включаются последовательно с элементами исходной системы, вторые — параллельно каким-либо элементам. Параллельные КУ являются местными обратными связями, которые могут быть как отрицательными, так и положительными.

Остановимся на одном, простом и эффективном способе определения математических моделей КУ и соответствующих этим моделям физических элементов, образующих корректирующие устройства. Способ базируется на использовании так называемых желаемых логарифмических амплитудно-частотных характеристик ($L_{\text{ж}}$). *Желаемой ЛАЧХ* называется такая характеристика $L(\omega)$, которая соответствует линейной динамической системе, полностью отвечающей требованиям, предъявляемым в динамических режимах. Можно сформулировать требования к форме $L_{\text{ж}}$ для широкого круга статических и астатических систем.

Наибольшее влияние на динамику оказывает форма $L_{\text{ж}}(\omega)$ в районе частоты среза $\omega_{\text{ср}}$. В связи с этим весь диапазон частот характеристики $L_{\text{ж}}(\omega)$ (рис. 2.12) разобьем на три части: участок низких частот ($0 < \omega < \omega_1$); участок средних частот ($\omega_2 < \omega < \omega_3$); высокочастотный участок ($\omega > \omega_3$).

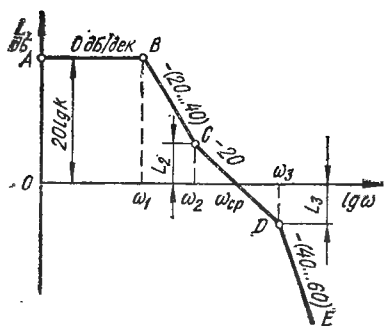


Рис. 2.12. Характеристика $L_{ж}(\omega)$

реализуемых физическими элементами, корректирующие устройства.

В практике проектирования линейных динамических систем широко используются зависимости, связывающие значения $\omega_{ср}$ с величинами $t_{\max}$ и $\delta_{\max}$, где $t_{\max}$ — максимальное время переходного процесса; $\delta_{\max}$ — максимальное перерегулирование. Используя эти зависимости, можно выбрать значение $\omega_{ср}$, обеспечивающее требуемые показатели качества управления в проектируемой системе.

Анализ систем, имеющих высокие динамические качества, показывает, что ЛАЧХ таких систем должна иметь в районе частоты среза наклон —20 дБ/дек. Следовательно, среднечастотный участок характеристики $L_{ж}(\omega)$ представляет собой линию CD (рис. 2.12), проходящую через точку $\omega_{ср}$ с наклоном —20 дБ/дек. Границы участка CD определяются ординатами L_2 и L_3 , соответствующими частотам ω_2 и ω_3 , которые также могут быть определены по заданным значениям $\delta_{\max}$.

Участок характеристики $L_{ж}(\omega)$ правее точки D проводится с таким наклоном, который имеет ЛАЧХ исходной системы в этом диапазоне частот, так как высокочастотный участок характеристики $L_{ж}(\omega)$ практически не влияет на показатели качества управления.

Низкочастотный участок $L_{ж}(\omega)$ для статических систем представляет собой прямую AB , параллельную оси частот и проходящую на высоте $20 \lg k$, где k — коэффициент усиления исходной разомкнутой системы. Для астатических систем этот участок характеристики представляет собой прямую линию с наклоном —20 ν дБ/дек., где ν — порядок астатизма, проходящую через точку с координатами $[\omega=1; L(\omega)=20 \lg k]$. Наконец, низкочастотный и среднечастотный участки ЛАЧХ сопрягаются отрезком BC (или более чем одним отрезком). Наклон отрезка BC выбирается из условия, чтобы перепад наклонов двух соседних отрезков характеристики $L_{ж}(\omega)$ не превышал 40 дБ/дек. Это требование вытекает из возможностей реальных физических элементов, с помощью которых реализуется характеристика $L_{ж}(\omega)$.

Для определения типа и параметров КУ на базе использования характеристики $L_{ж}(\omega)$ необходимо сравнить эту характеристику с ЛАЧХ $L_{исх}(\omega)$ исходной системы и установить вид ЛАЧХ требу-

Построение характеристики $L_{ж}(\omega)$ начинается с выбора значения частоты $\omega_{ср}$, которая прямо связана с быстродействием системы. Меньшие значения $\omega_{ср}$ соответствуют меньшему быстродействию системы, т. е. отражают большую длительность переходных процессов. При выборе значительных частот $\omega_{ср}$ для характеристики $L_{ж}(\omega)$ могут возникнуть непреодолимые трудности в практической реализации этой характеристики, так как требуются очень сложные, вплоть до не-

емого корректирующего устройства, т. е. форму характеристики $L_{к.у}(\omega)$.

Если синтезируется последовательное КУ, то

$$W_{ск}(p) = W_{исх}(p) W_{к.у}(p),$$

где $W_{ск}(p)$, $W_{исх}(p)$ — передаточные функции соответственно скорректированной и исходной систем; $W_{к.у}(p)$ — передаточная функция последовательного корректирующего устройства.

Переходя от передаточных функций к АФХ, а затем к ЛАЧХ, получим $L_{ск}(\omega) = L_{исх}(\omega) + L_{к.у}(\omega)$.

Так как цель коррекции сводится к тому, чтобы преобразовать исходную систему в желаемую с точки зрения динамических свойств, то очевидно, что $L_{ск}(\omega) = L_{ж}(\omega)$, следовательно, $L_{к.у}(\omega) = L_{ж}(\omega) - L_{исх}(\omega)$. Таким образом, если из ординат характеристики $L_{ж}(\omega)$ вычесть ординаты характеристики $L_{исх}(\omega)$, то получим ординаты ЛАЧХ последовательного КУ. Далее по виду $L_{к.у}(\omega)$ можно выбрать физические элементы, реализующие эту ЛАЧХ, т. е. выбрать физическую схему и параметры КУ.

При выборе параллельного КУ

$$W_{ск}(p) = \frac{W_{охв}(p) W_{н.о}(p)}{1 + W_{охв}(p) W_{к.у}(p)},$$

где $W_{охв}(p)$, $W_{н.о}(p)$ — передаточные функции части исходной системы соответственно охваченной местной обратной связью (параллельным КУ) и не охваченной ею.

Переходя к АФХ, получим

$$W_{ск}(j\omega) = \frac{W_{охв}(j\omega) W_{н.о}(j\omega)}{1 + W_{охв}(j\omega) W_{к.у}(j\omega)}.$$

Для диапазона частот, в котором $|W_{охв}(j\omega) W_{н.о}(j\omega)| \geq 1$, имеем $W_{ск}(j\omega) \approx W_{н.о}(j\omega) / W_{к.у}(j\omega)$.

Переходя к ЛАЧХ, получим

$$L_{к.у}(\omega) = L_{н.о}(\omega) - L_{ск}(\omega),$$

т. е. ЛАЧХ параллельного корректирующего устройства имеет вид

$$L_{к.у}(\omega) = L_{н.о}(\omega) - L_{исх}(\omega).$$

На основе последнего выражения можно определить логарифмическую частотную характеристику параллельного КУ, выбрать физические элементы и их параметры, реализующие $L_{к.у}(\omega)$. В практике проектирования линейных динамических систем широко используются таблицы, содержащие физические схемы КУ, их передаточные функции и характеристики $L_{к.у}(\omega)$. В таких случаях задача выбора КУ сводится к определению вышеуказанным способом формы $L_{к.у}(\omega)$ и выбору в таблицах необходимого типа корректирующего устройства.

Целесообразно подчеркнуть, что решение задачи выбора КУ неоднозначно, так как одни и те же динамические характеристики можно обеспечить различными физическими элементами, образующими корректирующие устройства. Например, в практике автоматического управления широко используются параллельные

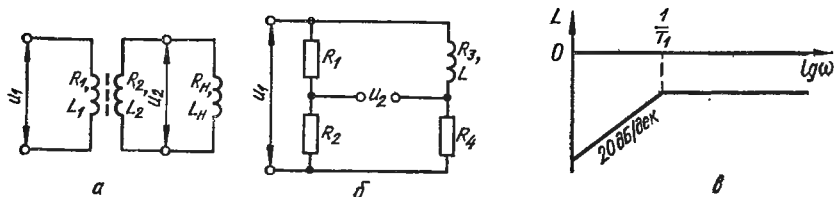


Рис. 2.13. Корректирующие устройства, реализующие местную гибкую обратную связь:

а — стабилизирующий трансформатор; б — динамический мост; в — ЛАХ

КУ, реализующие гибкую обратную связь по различным физическим переменным, подаваемым на вход КУ, с выхода охваченной части линейной системы. Такие КУ имеют передаточную функцию вида

$$W(p) = T_2(p)/(T_1p + 1).$$

Эту передаточную функцию и соответствующую ей $L_{к.у}(\omega)$ можно реализовать с помощью нескольких устройств, часть из которых приведена на рис. 2.13. Постоянные времени и коэффициенты усиления их определяются так:

для стабилизирующего трансформатора

$$T_1 = L_1/R_1; \quad T_2 = T_1 \frac{k_{тр}}{R_2 + R_H}; \quad k_{тр} = \frac{u_1}{u_2};$$

для динамического моста

$$T_1 = L/R_3; \quad T_2 = T_1 k_M; \\ k_M = \frac{R_2}{R_1 + R_2} = \frac{R_4}{R_3 + R_4}.$$

Глава 3

НЕЛИНЕЙНЫЕ ДИНАМИЧЕСКИЕ СИСТЕМЫ УПРАВЛЕНИЯ

3.1. ОПРЕДЕЛЕНИЕ И ОСОБЕННОСТИ НЕЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ УПРАВЛЕНИЯ

Нелинейными называются системы, содержащие нелинейные элементы (НЭ), т. е. такие элементы, у которых соотношения между выходными и входными величинами описываются нелинейными уравнениями. Строго говоря, все динамические системы являются нелинейными. В этом смысле линейные методы анализа и синтеза динамических систем управления неприемлемы. Однако линейные модели реальных нелинейных систем во многих случаях могут с достаточным приближением отражать свойства реальных систем. Кроме того, методы исследования нелинейных систем часто базируются на методах линейной теории с добавлением тех или иных особенностей, учитывающих существование нелинейностей. В целом же тео-

рия нелинейных систем является более сложной, менее разработанной. Это, в частности, объясняется большим разнообразием типов нелинейных систем, что затрудняет разработку универсальных методов для их исследования. Тем не менее, применение существующих методов исследования нелинейных систем во многих случаях дает достаточно эффективный результат.

Широкий круг нелинейных динамических систем содержит нелинейные элементы с типовыми нелинейностями, под которыми понимаются определенные статические зависимости $x_{\text{вых}} = f(x_{\text{вх}})$. Некоторые из этих зависимостей приведены в табл. 3.1. Нелинейные элементы, у которых зависимость $x_{\text{вых}} = f(x_{\text{вх}})$ одинакова как при возрастании величины $x_{\text{вх}}$, так и при ее убывании, называются *нелинейными элементами с однозначными статическими характеристиками*. Нелинейные элементы, у которых при возрастании $x_{\text{вх}}$ имеет место одна зависимость $x_{\text{вых}} = f(x_{\text{вх}})$, а при уменьшении $x_{\text{вх}}$ — другая, называются *нелинейными элементами с неоднозначными статическими характеристиками*.

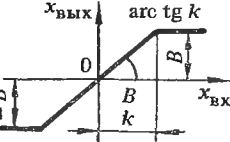
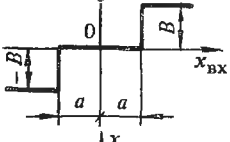
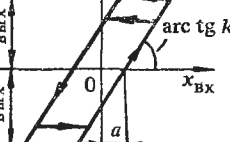
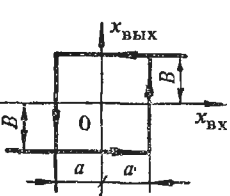
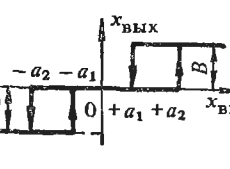
Неоднозначные статические характеристики отражают наличие гистерезисных явлений в соответствующих устройствах автоматики.

Нелинейные элементы обуславливают существенные отличия свойств нелинейных систем от линейных, что требует разработки иных методов исследования. Характер возможных движений в нелинейных системах более широк, чем в линейных. Действительно, в линейной системе движение может быть либо сходящимся, если система устойчива, либо расходящимся, если система неустойчива. Существование устойчивых колебательных движений в линейных системах недопустимо, так как означает нахождение системы на границе устойчивости. Такой режим работы линейной системы практически невозможен, поскольку при любом незначительном изменении ее параметров движения координат системы станут либо сходящимися либо расходящимися. В нелинейной же системе возможны устойчивые колебательные движения тех или иных переменных при отсутствии внешних периодических возмущений. Такие движения называются *автоколебаниями*, а системы — *автоколебательными*. Амплитуда и частота автоколебаний определяются параметрами системы. Следовательно, в одной и той же нелинейной системе могут быть несколько режимов автоколебаний.

Автоколебания могут быть нежелательными или недопустимыми. В таком случае параметры системы выбираются из условий невозможности автоколебательных режимов. Однако автоколебания могут быть необходимыми, обеспечивая принцип работы конкретной динамической системы, например вибратора-уплотнителя какой-либо массы. В этом случае параметры системы выбираются из условий обеспечения требуемых устойчивых автоколебательных режимов.

Природа возникновения автоколебаний связана с тем, что коэффициент усиления нелинейной системы на разных участках статической характеристики различен, что вызывает автоматический переход от расходящихся движений к сходящимся и наоборот.

Таблица 3.1

Графическая зависимость	Тип нелинейности	Аналитическая зависимость
	Нечувстви- тельность	$x_{\text{вых}} = 0 \text{ при } x_{\text{вх}} \leq a$ $x_{\text{вых}} = -ka + kx_{\text{вх}} \text{ при } x_{\text{вх}} \geq a$ $x_{\text{вых}} = ka - kx_{\text{вх}} \text{ при } x_{\text{вх}} < -a$
	Ограничение (насыщение)	$x_{\text{вых}} = kx_{\text{вх}} \text{ при } x_{\text{вх}} \leq \frac{B}{k}$ $x_{\text{вых}} = B \text{ при } x_{\text{вх}} \geq \frac{B}{k}$ $x_{\text{вых}} = -B \text{ при } x_{\text{вх}} \leq -\frac{B}{k}$
	Релейная	$x_{\text{вых}} = B \text{ при } x_{\text{вх}} > 0$ $x_{\text{вых}} = -B \text{ при } x_{\text{вх}} < 0$
	Релейная с зоной нечувстви- тельности	$x_{\text{вых}} = B \text{ при } x_{\text{вх}} \geq a$ $x_{\text{вых}} = 0 \text{ при } -a \leq x_{\text{вх}} \leq -a$ $x_{\text{вых}} = -B \text{ при } x_{\text{вх}} \leq -a$
	Зазор (гистерезис)	$x_{\text{вых}} = -ka + kx_{\text{вх}} \text{ при } x' > a$ $x_{\text{вых}} = ka - kx_{\text{вх}} \text{ при } x' < -a$ $\frac{dx_{\text{вых}}}{dx_{\text{вх}}} = 0 \text{ при } \frac{x_{\text{вых}}}{k} - x_{\text{вх}} < a$
	Релейная с гистерезисом	$\left. \begin{aligned} x_{\text{вых}} &= B \text{ при } x_{\text{вх}} \geq a \\ x_{\text{вых}} &= -B \text{ при } x_{\text{вх}} \leq -a \end{aligned} \right\} x'_{\text{вх}} > 0$ $\left. \begin{aligned} x_{\text{вых}} &= B \text{ при } x_{\text{вх}} \geq -a \\ x_{\text{вых}} &= -B \text{ при } x_{\text{вх}} \leq -a \end{aligned} \right\} x'_{\text{вх}} < 0$
	Релейная с нечувстви- тельностью и гистерезисом	$\left. \begin{aligned} x_{\text{вых}} &= B \text{ при } x_{\text{вх}} \geq a_2 \\ x_{\text{вых}} &= 0 \text{ при } -a_1 \leq x_{\text{вх}} \leq a_2 \\ x_{\text{вых}} &= -B \text{ при } x_{\text{вх}} \leq -a_1 \end{aligned} \right\} x'_{\text{вх}} > 0$ $\left. \begin{aligned} x_{\text{вых}} &= B \text{ при } x_{\text{вх}} \geq a_1 \\ x_{\text{вых}} &= 0 \text{ при } -a_2 \leq x_{\text{вх}} \leq a_1 \\ x_{\text{вых}} &= -B \text{ при } x_{\text{вх}} \leq -a_2 \end{aligned} \right\} x'_{\text{вх}} < 0$

В нелинейных системах существенно изменяется постановка и решение задачи анализа устойчивости. Как было показано в предыдущей главе, устойчивость линейной системы зависит только от свойств системы и не зависит от отклонения переменных в начальный момент. В нелинейной же системе, устойчивой при малых отклонениях, могут возникать неустойчивые режимы при больших начальных отклонениях. В связи с этим для нелинейной системы вводятся понятия «устойчивость в малом» и «устойчивость в большом».

Широкий класс нелинейных динамических систем может быть описан уравнением вида

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n), \quad (3.1)$$

где x_i — координаты системы; $f_i, i = \overline{1, n}$ — непрерывные или кусочно-непрерывные функции.

Сущность уравнений типа (3.1) состоит в том, что под знаком нелинейной функции стоят только сами переменные, без производных, а исходная система уравнений, описывающих нелинейную систему, решена относительно первых производных всех координат по времени. К числу систем, описываемых уравнением вида (3.1), относятся системы с выделенным нелинейным элементом, например элементом с типовой нелинейной статической характеристикой. Такая система может быть представлена как последовательное соединение линейной части (ЛЧ) и нелинейного элемента (НЭ), что в значительной степени упрощает исследование системы.

Точные методы анализа нелинейных динамических систем применимы лишь для ограниченного круга систем, описываемых уравнениями не выше второго-третьего порядка. Поэтому в теории нелинейных систем широко используются различные приближенные методы: численные, графические, графоаналитические. Использование численных методов связано с необходимостью выполнения громоздких вычислений, а графические методы затрудняют анализ влияния отдельных параметров на динамические свойства нелинейных систем.

3.2. ИССЛЕДОВАНИЕ НЕЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ МЕТОДОМ ФАЗОВЫХ ТРАЕКТОРИЙ

Метод фазовых траекторий основан на том, что состояние динамической системы характеризуется положением некоторой точки в плоскости (или пространстве) координат, характеризующей систему. Точка в пространстве координат системы, характеризующая состояние системы в любой момент времени, называется *изображающей точкой*. Ее проекция на оси координат в каждый момент времени дает значение той или иной переменной, характеризующей систему. Переменные оси координат выбираются так, чтобы изображающая точка давала наиболее полную информацию о состоянии системы. Например, если процесс рассматривается на плоскости,

т. е. в системе двух осей, координатными осями могут быть величины $x_{\text{ВЫХ}}$ и $x_{\text{ВХ}}$ или $x_{\text{ВХ}}$ и $x_{\text{ВХ}}$ и т. д. Из сказанного ясно, что метод фазовых траекторий дает наиболее простые и наглядные результаты при исследовании систем не выше второго-третьего порядка.

Основные идеи метода фазовых траекторий удобно рассмотреть применительно к линейным системам второго порядка, распространив потом результаты на нелинейные системы. Исходя из этого, обратимся к динамической системе, описываемой линейным дифференциальным уравнением второго порядка или двумя линейными дифференциальными уравнениями первого порядка:

$$\frac{dx_1}{dt} = ax_1 + bx_2; \quad \frac{dx_2}{dt} = cx_1 + ex_2, \quad (3.2)$$

где a, b, c, e — постоянные коэффициенты.

Решить систему уравнений (3.2) — это значит определить переменные x_1 и x_2 как функции времени и начальных условий x_{10}, x_{20} , т. е. определить $x_1 = f_1(t; x_{10}; x_{20})$; $x_2 = f_2(t; x_{10}; x_{20})$. Удобнее, однако, решение системы уравнений (3.2) представить в плоскости x_1, x_2 , называемой *фазовой плоскостью*, рассматривая время t как параметр. При таком подходе состояние системы в каждый момент времени характеризуется положением изображающей точки c_i на плоскости переменных x_1, x_2 (рис. 3.1). Начальное положение точки c_i определяется значением начальных условий x_{10}, x_{20} . Далее, с течением времени изображающая точка движется по некоторой траектории, определяемой свойствами динамической системы. Геометрическое место изображающих точек на фазовой плоскости (или в фазовом пространстве, если координат более двух) называется *фазовой траекторией*. Каждому значению начальных условий соответ-

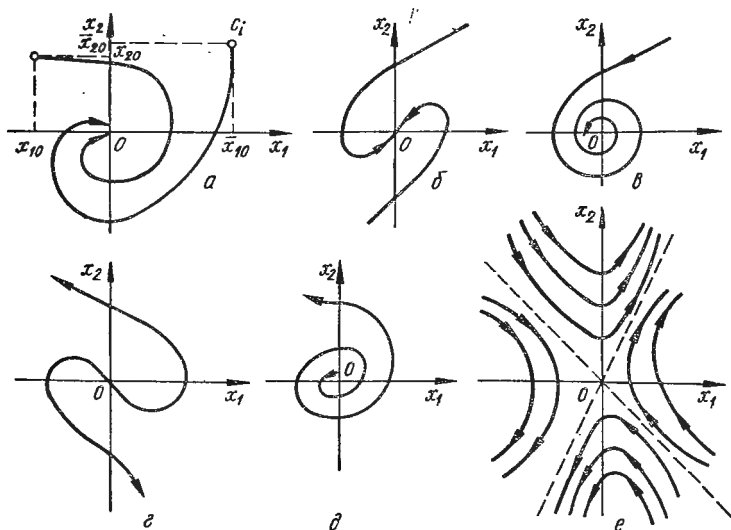


Рис. 3.1. Фазовые траектории и особые точки линейных систем:

а — траектория; б — особая точка типа «устойчивый узел»; в — «устойчивый фокус»; г — «неустойчивый узел»; д — «неустойчивый фокус»; е — «седло»

ствуется своя фазовая траектория. Вся совокупность фазовых траекторий на фазовой плоскости (или в фазовом пространстве) называется *фазовым портретом системы управления*.

Уравнение фазовых траекторий получим, поделив второе уравнение (3.2) на первое и исключив таким образом время:

$$\frac{dx_2}{dx_1} = \frac{cx_1 + ex_2}{ax_1 + bx_2}. \quad (3.3)$$

Очевидно, что в каждой точке фазовой плоскости выражение (3.3) определяет положение касательной к фазовой траектории, соответствующей заданным начальным условиям. Из (3.3) видно, что любой паре значений x_1 и x_2 соответствует одно единственное

значение $\frac{dx_2}{dx_1}$, т. е. в каждой точке фазовой траектории можно провести лишь одну касательную, а это означает, что фазовые траектории линейной системы на фазовой плоскости не пересекаются. Исключение составляет случай, когда $x_1 = x_2 = 0$ — точка начала координат фазовой плоскости. В этом случае, как следует из (3.3), $dx_2 = dx_1 = 0$ и положение касательной становится неопределенным: в точке можно провести более одной касательной, а следовательно, фазовые траектории в такой точке пересекаются. Точка на фазовой плоскости, в которой фазовые траектории пересекаются, называется *особой точкой*. Таким образом, на фазовой плоскости линейной динамической системы имеется лишь одна особая точка, которая находится в начале координат.

В зависимости от характера движения изображающей точки к особой точке особые точки получают то или иное название (рис. 3.1). Очевидно, представленные на рис. 3.1, а, б, в формы фазовых характеристик соответствуют устойчивым системам, а на рис. 3.1, г, д, е — неустойчивым.

Таким образом, фазовые портреты линейных динамических систем полностью отражают характер возможных движений переменных, описывающих эти системы.

Распространим категории и понятия аппарата фазовых траекторий на нелинейные динамические системы, ограничившись системами второго порядка. С этой целью обратимся к нелинейной системе, описываемой двумя нелинейными уравнениями первого порядка:

$$\frac{dx_1}{dt} = F_1(x_1, x_2); \quad \frac{dx_2}{dt} = F_2(x_1, x_2), \quad (3.4)$$

где $F_1(x_1, x_2)$ и $F_2(x_1, x_2)$ — нелинейные функции.

Так же, как и в линейных системах, уравнение фазовых траекторий получим, поделив второе уравнение (3.4) на первое, исключив тем самым время:

$$\frac{dx_2}{dx_1} = \frac{F_2(x_1, x_2)}{F_1(x_1, x_2)}. \quad (3.5)$$

Особые точки в фазовом пространстве рассматриваемой нелинейной системы определяются как общие корни двух уравнений:

$$F_1(x_1, x_2) = 0; \quad F_2(x_1, x_2) = 0. \quad (3.6)$$

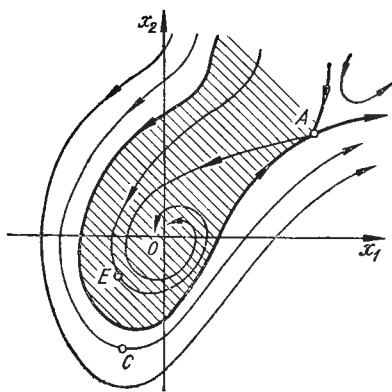


Рис. 3.2. Фазовые траектории нелинейной динамической системы

При этом одна особая точка находится в начале координат, где $x_1 = x_2 = 0$ и $\frac{dx_2}{dx_1} = 0$. В линейных системах функции (3.6) представляют собой прямые линии, которые, пересекаясь в начале координат, естественно, больше в фазовой плоскости не пересекаются. В нелинейных же системах, пересекаясь в начале координат, функции (3.6) в силу нелинейности могут пересекаться также в других точках фазовой плоскости. Следовательно, на фазовой плоскости нелинейной динамической системы может быть несколько особых точек, а значит, не-

сколько равновесных состояний системы. Особые точки имеют те же названия, что и на фазовых плоскостях линейных систем: узлы, фокусы, седла. На рис. 3.2 в качестве примера приведен фазовый портрет нелинейной системы с двумя особыми точками, одна из которых — устойчивый фокус — находится в начале координат (точка 0), другая — седло (точка A). Как видно из рисунка, характер движения в рассматриваемой системе зависит от начального отклонения ее координат: если начальные отклонения находятся в пределах заштрихованной зоны, движение в системе носит сходящийся характер, а если за ее пределами — движения расходящиеся.

Как отмечалось выше, в нелинейных динамических системах возможны устойчивые автоколебания, т. е. значение переменных, характеризующих систему, может периодически повторяться. Это свойство приводит к возможности появления замкнутых траекторий на фазовых портретах нелинейных систем, которые называются *предельными циклами* (рис. 3.3). Предельные циклы могут быть устойчивыми и неустойчивыми. В первом случае изображающая точка из любой точки фазовой плоскости стремится попасть на замкнутую траекторию, во втором — уйти от замкнутой траектории либо к началу координат, либо в бесконечность.

Для выяснения устойчивости или неустойчивости предельного

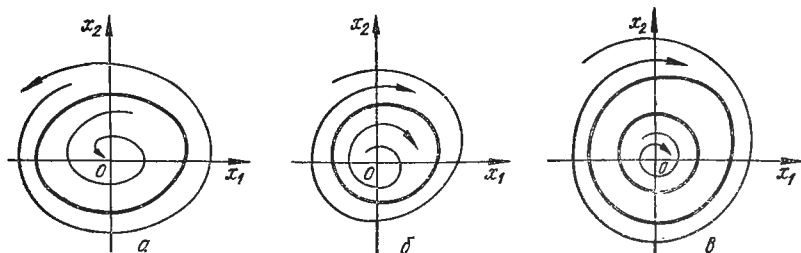


Рис. 3.3. Предельные циклы фазовых траекторий нелинейных систем

цикла необходимо построить две фазовые траектории: одну — внутри предельного цикла, другую — за его пределами и проследить движение изображающей точки по этим траекториям. Предельный цикл, показанный на рис. 3.3, *а*, является неустойчивым, а сама нелинейная система, имеющая такой фазовый портрет, устойчива в малом и неустойчива в большом. Действительно, если переходный процесс начинается при начальных отклонениях координат x_1 и x_2 , лежащих внутри предельного цикла, то он будет сходящимся. Если начальные отклонения координат находятся за пределами замкнутой траектории, то переходный процесс будет расходящимся.

Рассуждая аналогично, можно убедиться в том, что предельный цикл, представленный на рис. 3.3, *б*, является устойчивым. Фазовый портрет (рис. 3.3, *в*) содержит два предельных цикла, из которых внешний является устойчивым, а внутренний — неустойчивым. Таким образом, предельные циклы на фазовых портретах нелинейных систем могут играть роль границы устойчивости. При наличии нескольких особых точек области устойчивости выделяют незамкнутые кривые, которые называются *сепаратрисами*. Предельные циклы, особые точки, сепаратрисы определяют топологию фазового портрета конкретной нелинейной системы. При изменениях параметров нелинейной динамической системы в определенных пределах структура фазового портрета может оставаться неизменной. Однако существуют некоторые дискретные значения параметров системы, при которых топологическая структура фазового портрета претерпевает изменения. Эти дискретные значения параметров системы называются *бифуркационными*.

Полный анализ нелинейной динамической системы включает определение возможных в системе фазовых портретов, бифуркационных границ в пространстве параметров, формы и расположения предельных циклов и сепаратрис по отношению к каждой точке пространства параметров. Однако такой полный анализ нелинейной системы, как правило, выполнить не удастся из-за чрезвычайной сложности задачи.

Наиболее часто в инженерной практике встречается задача построения фазовых портретов нелинейных систем, в которых можно выделить нелинейную часть и типовую нелинейность. В этом случае поведение нелинейной системы удастся описать совокупностью линейных дифференциальных уравнений, каждое из которых справедливо в пределах соответствующих линейных отрезков типовой нелинейной статической характеристики. После того как определены фазовые траектории, соответствующие линейным дифференциальным уравнениям, на фазовую плоскость наносятся «линии переключения», которые представляют собой геометрические места точек фазовой плоскости, где осуществляется переход от фазовых траекторий, соответствующих одному линейному уравнению, к фазовой траектории, обусловленной другим линейным уравнением.

В качестве примера рассмотрим нелинейную систему регулирования температуры объекта управления ОУ (рис. 3.4), например температуры пропарочной камеры.

Температура Θ объекта (рис. 3.4, *а*) контролируется с помощью электриче-

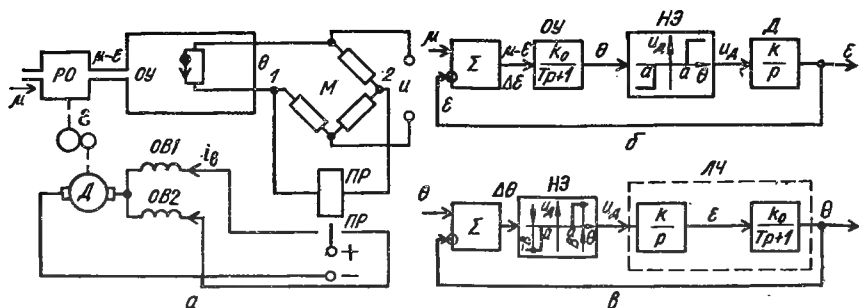


Рис. 3.4. Нелинейная система автоматического регулирования температуры:
а — функциональная схема; б, в — структурные схемы

ской мостовой схемы М, в одно плечо которой включен датчик температуры. При заданном значении Θ мост уравновешен и напряжение, подводимое к обмотке поляризованного реле ПР, равно нулю. При этих условиях контакт реле ПР занимает среднее положение. При отклонении температуры Θ от заданного значения контакт ПР замкнет цепь электродвигателя Д на то или иное направление вращения. Двигатель Д через зубчатую передачу воздействует на регулирующий орган РО путем изменения координаты ϵ , что вызовет изменение подачи теплоносителя μ к объекту управления. Направление вращения двигателя определяется тем, какая из его обмоток возбуждения — ОВ₁ или ОВ₂ — включена, что зависит от положения контакта ПР, а следовательно, от знака отклонения Θ .

Уравнение объекта управления рассматриваемой системы

$$T \frac{d\Theta}{dt} + \Theta = k_0 \Delta \epsilon \quad \text{или} \quad \Theta(p)(Tp + 1) = k_0 \Delta \epsilon(p),$$

где T — постоянная времени; k_0 — коэффициент усиления объекта.

Уравнение электродвигателя

$$\epsilon(p)/u_d(p) = k_n/p.$$

Структурная схема, соответствующая написанным уравнениям динамики с учетом характера нелинейности, приведена на рис. 3.4, б.

Объединяя линейные звенья в одну линейную часть (ЛЧ), получим схему, представленную на рис. 3.4, в, откуда видно, что уравнение линейной части системы

$$p(Tp + 1)\Theta(p) = ku_d(p), \quad (3.7)$$

где $k = k_0 k_n$ — коэффициент усиления линейной части.

Для общности обозначим $\Theta = x$, а напряжение u_d после включения ПР обозначим $u_d = \pm B$. Тогда, в зависимости от значения u_d , состояние рассматриваемой системы регулирования описывается одним из следующих трех уравнений:

$$T \frac{d^2x}{dt^2} + \frac{dx}{dt} = -kB \quad \text{при} \quad u_d = -B; \quad (3.8)$$

$$T \frac{d^2x}{dt^2} + \frac{dx}{dt} = 0 \quad \text{при} \quad u_d = 0; \quad (3.9)$$

$$T \frac{d^2x}{dt^2} + \frac{dx}{dt} = kB \quad \text{при} \quad u_d = B. \quad (3.10)$$

Для построения фазового портрета выберем координаты фазовой плоскости

x и $y = x'$. Тогда (3.9) принимает вид $Ty' + y = 0$; или $T \frac{dy}{dx} y + y = 0$, т. е. $T \frac{dy}{dx} + 1 = 0$, откуда

$$y = -x/T + c. \quad (3.11)$$

Уравнение (3.11) представляет собой уравнение фазовых траекторий рассматриваемой системы автоматического регулирования температуры в режимах, описываемых уравнением (3.9). Как видно из (3.11), фазовые траектории на этом участке представляют собой параллельные прямые линии, отстоящие друг от друга на величину ординаты c , равной постоянной интегрирования (рис. 3.5). Ширина рассматриваемого участка равна $2a$ — зоне нечувствительности поляризованного реле, т. е. зоне нечувствительности нелинейного элемента.

Аналогично, получим уравнение фазовых траекторий:

для режима работы, соответствующего условию (3.8),

$$x = T[kB \ln(y + kB) - y] + c_1; \quad (3.12)$$

для режима работы, соответствующего условию (3.10),

$$x = -T[kB \ln(y - kB) + y] + c_2. \quad (3.13)$$

Фазовые траектории, построенные по уравнениям (3.11)–(3.13), представлены на рис. 3.5.

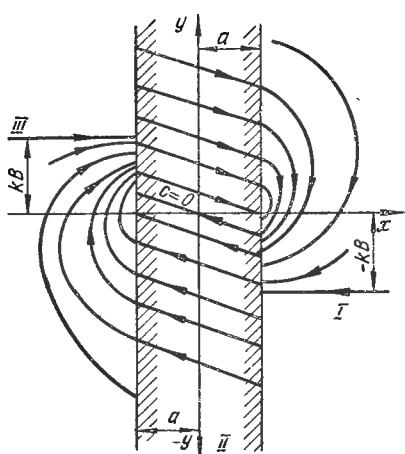


Рис. 3.5. Фазовые портреты нелинейной системы с релейной типовой нелинейностью с зоной нечувствительности (I — уравнение (3.12), II — (3.11), III — (3.13))

3.3. МЕТОД ГАРМОНИЧЕСКОГО БАЛАНСА

Метод гармонического баланса как один из методов исследования нелинейных динамических систем основан на использовании свойства реальных элементов систем не пропускать колебания высших гармоник — свойства фильтра. Фильтрующие свойства элементов динамических систем зависят от их инерционности. В тех случаях, когда эти свойства проявляются достаточно ярко, периодические режимы в системе носят характер, близкий к синусоидальному закону. Тогда, независимо от характера периодических возмущений, приложенных к системе, на ее выходе можно рассматривать лишь колебания основной гармонической составляющей.

Использование метода гармонического баланса позволяет сравнительно легко исследовать автоколебательные режимы в нелинейных системах на устойчивость, определить частоту и амплитуду устойчивых автоколебаний с достаточной степенью приближения. Рассмотрим основы метода на примере нелинейной системы, состоящей из линейной части и безынерционного нелинейного элемента НЭ (рис. 3.6).

В режиме автоколебаний каждая из координат системы совершает периодическое движение с одинаковой частотой, но с разными амплитудами и фазами. На основе сформулированного выше прин-

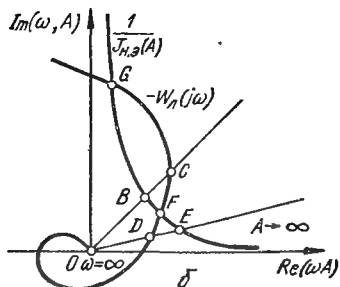
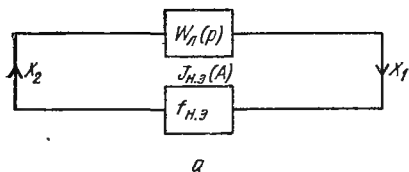


Рис. 3.6. К методу гармонического баланса:
а — нелинейная система; б — ее динамические характеристики

ципа гармонического баланса считаем колебания координат системы синусоидальными. В этом случае для характеристики динамических свойств линейной части системы можно использовать амплитудно-фазовую характеристику линейной части:

$$W_n(j\omega) = X_1/X_2. \quad (3.14)$$

Свойства нелинейного элемента характеризуются формой его статической характеристики $X_2 = f_{н.э}(X_1)$. Для оценки динамических свойств НЭ введем также понятие эквивалентной амплитудно-фазовой характеристики нелинейного элемента $J_{н.э}(A)$, где A — отношение амплитуды величины X_2 на выходе безынерционного НЭ к амплитуде синусоидально изменяющейся величины X_1 на входе этого элемента. В силу безынерционности НЭ характеристика $J_{н.э}(A)$, являясь функцией амплитуды величины на выходе НЭ, не зависит от частоты изменения входной и выходной величин.

Изложенные предпосылки — принцип гармонического баланса и безынерционность нелинейного элемента — позволяют считать, что характеристика $J_{н.э}(A)$ полностью определяется формой статической характеристики и может быть определена по заданной зависимости $X_2 = f_{н.э}(X_1)$. Действительно, в соответствии с принципом фильтра в режиме автоколебаний на входе НЭ имеется синусоидально изменяющаяся величина с амплитудой A и частотой ω :

$$X_1 = A \sin \omega t.$$

Величина X_2 на выходе НЭ в общем случае будет изменяться по некоторому периодическому закону, определяемому свойствами НЭ. Если нелинейный элемент обладает достаточными фильтрующими свойствами, то на его выходе можно принимать во внимание только первую гармоническую составляющую разложения в ряд Фурье величины $X_2(t)$, исключив из рассмотрения составляющие, содержащие высшие гармоники. Исходя из этого, величина на выходе НЭ определится так:

$$X_2(t) = f_{н.э}[X_1(t)] = f_{н.э}[A \sin \omega t]$$

или

$$X_2(t) \approx A[B(A) \sin \omega t + C(A) \cos \omega t],$$

Графическая зависимость	Значения коэффициентов
	$B(A) = \frac{4f_0}{\pi A}; \quad C(A) = 0$
	$B(A) = \frac{4f_0}{\pi A} \sqrt{1 - \left(\frac{a}{A}\right)^2}; \quad C(A) = 0$
	$B(A) = \alpha - \frac{2}{\pi} \left[\arcsin \frac{a}{A} \left[\arcsin \frac{a}{A} + \frac{a}{A} \sqrt{1 - \left(\frac{a}{A}\right)^2} \right] \right];$ $C(A) = 0$
	$B(A) = \alpha_2 + \frac{2}{\pi} (\alpha_1 - \alpha_2) \left[\arcsin \frac{a}{A} + \frac{a}{A} \sqrt{1 - \left(\frac{a}{A}\right)^2} \right]$
	Для $A > a + \frac{f_0}{2}$ $B(A) = \frac{2\alpha}{\pi} (\alpha_2 - \alpha_1 + \sin 2\alpha_2 - \sin \alpha_1)$ где $\arcsin \frac{a}{A} = \sigma_1 < \frac{\pi}{2}$ $\sigma_2 = \arcsin \frac{1}{A} \left(a + \frac{f_0}{\alpha} \right) < \frac{\pi}{2} \quad C(A) = 0$
	$B(A) = \frac{4f_0}{\pi A} \sqrt{1 - \left(\frac{a}{A}\right)^2}; \quad C(A) = \frac{4f_0 a}{\pi A^2}$
	$B(A) = \frac{1}{\pi} (\psi_1 - \frac{1}{2} \sin 2\psi_1); \quad C(A) = \frac{1}{\pi} \sin \psi_1$ где $\psi_1 = \arccos \left[1 - \frac{2(A-a)}{A} \right] < \pi$

где $B(A)$, $C(A)$ — коэффициенты первой гармоники разложения функции $X_2(t)$ в ряд Фурье.

При разложении функции $X_2(t)$ в ряд Фурье не учтена постоянная составляющая разложения, так как предметом рассмотрения являются ординаты переменных, обусловленные автоколебаниями.

Коэффициенты $B(A)$ и $C(A)$ определяются как коэффициенты ряда Фурье:

$$B(A) = \frac{1}{\pi A} \int_0^{2\pi} f_{н.э}(A \sin \omega t) \sin \omega t d\omega t;$$

$$C(A) = \frac{1}{\pi A} \int_0^{2\pi} f_{н.э}(A \sin \omega t) \cos \omega t d\omega t.$$

Имея значения коэффициентов $B(A)$ и $C(A)$, можно найти характеристику

$$J_{н.э}(A) = B(A) + jC(A).$$

Коэффициенты $B(A)$ и $C(A)$ могут быть вычислены при известных зависимостях $X_2 = f_{н.э}(X_1)$. Для ряда типовых нелинейностей коэффициенты $B(A)$ и $C(A)$ приведены в табл. 3.2. Для однозначных нелинейных характеристик $C(A) = 0$.

Проведенные рассуждения позволяют приближенно заместить нелинейный элемент эквивалентным линейным элементом с характеристикой $J_{н.э}(A)$, у которого отношение модулей векторов синусоидально изменяющихся выходной и входной величин такое же, как у исходного нелинейного элемента (рис. 3.6, а):

$$J_{н.э}(A) = X_2/X_1. \quad (3.15)$$

Перемножая уравнение (3.14) и (3.15), получаем

$$W_{л}(j\omega) J_{н.э}(A) = 1$$

или с учетом замыкания системы отрицательной обратной связью

$$-W_{л}(j\omega) = 1/J_{н.э}(A). \quad (3.16)$$

Выражение (3.16) можно рассматривать как условие возникновения периодических движений с частотой ω и амплитудой A . Из (3.16) следует, что в точках пересечения характеристик $-W_1(j\omega)$ и $1/J_{н.э}(A)$ (рис. 3.6, б) можно определить параметры автоколебаний. В этих точках по характеристике $-W_{л}(j\omega)$ определяется частота автоколебаний, а по характеристике $1/J_{н.э}(A)$ — амплитуда.

Из (3.16) следует, что

$$-W_{л}(j\omega) J_{н.э}(A) = 1. \quad (3.17)$$

Таким образом, значение модуля произведения (3.17), равное единице, означает наличие автоколебательного режима. Очевидно, что автоколебания будут расходящимися в точках комплексной плоскости (рис. 3.6, б), для которых имеет место условие

$|W_{л}(j\omega) J_{н.э}(A)| > 1$, и сходящимися при условии $|W_{л}(j\omega) J_{н.э}(A)| < 1$. Сформулированные условия позволяют проанализировать устойчивость автоколебательных режимов и прийти к выводу, что автоколебания с параметрами, определяемыми точкой F , т. е. с частотой ω_F и амплитудой A_F , устойчивы, а автоколебания с параметрами, определяемыми точкой G , т. е. с частотой ω_G и амплитудой A_G , — неустойчивы.

Действительно, рассмотрим два динамических режима, соседних с режимом, определяемым точкой F . Система, выведенная из состояния устойчивых автоколебаний в сторону увеличения параметров автоколебаний (частоты и амплитуды), будет стремиться вернуться в равновесное состояние, т. е. в состояние, определяемое точкой F . Для режима, характеризуемого точками D и E , имеем

$$A_E > A_F; \quad \omega_D > \omega_F;$$

$$1/J_{H.3}(A_E) = \overline{0E}; \quad -W_{\Pi}(j\omega) = \overline{0D}; \quad |W_{\Pi}(j\omega_D)J_{H.3}(A_E)| = \overline{0D}/\overline{0E}.$$

Так как $\overline{0D} < \overline{0E}$, то $|W_{\Pi}(j\omega_D)J_{H.3}(A_E)| < 1$, т. е. режим автоколебаний с параметрами $\omega_D > \omega_F$ и $A_E > A_F$ будет сходящимся, и система стремится возвратиться в состояние, характеризуемое точкой F .

Для режима $A_B < A_F$ и $\omega_C < \omega_F$ (точки B и C) имеем:

$$|W_{\Pi}(j\omega_C)J_{H.3}(A_B)| = \overline{0C}/\overline{0B}.$$

Как как $\overline{0C} > \overline{0B}$, то $|W_{\Pi}(j\omega_C)J_{H.3}(A_B)| > 1$, т. е. режим автоколебаний с параметрами $A_B < A_F$ и $\omega_C < \omega_F$ будет расходящимся, и система вновь стремится в состояние, характеризуемое точкой F .

Проведя аналогичные рассуждения, можно убедиться в том, что выведенная из режима, характеризуемого точкой G , система не стремится возвратиться в этот режим, следовательно, автоколебания этого режима неустойчивы. Полученные результаты позволяют сформулировать критерий устойчивости автоколебаний, базирующийся на принципах гармонического баланса, применительно к нелинейной системе, состоящей из линейной части и одного безынерционного нелинейного элемента: режим автоколебаний будет устойчивым, если годограф амплитудно-фазовой характеристики линейной части системы, взятой с обратным знаком $[-W_{\Pi}(j\omega)]$, не охватывает точку годографа обратной эквивалентной амплитудно-фазовой характеристики $1/J_{H.3}(A)$ нелинейного элемента, соответствующую несколько увеличенной амплитуде по сравнению с амплитудой в точке пересечения характеристик $-W_{\Pi}(j\omega)$ и $1/J_{H.3}(A)$.

В тех случаях, когда под знак нелинейной функции входят не только координаты системы, но и их производные, более удобным для анализа нелинейной системы становится метод гармонического баланса, разработанный Е. П. Поповым. Суть метода в том, что нелинейное уравнение приводится к формально линейному, но содержащему коэффициенты, зависящие от частоты и амплитуды колебаний. В результате для фиксированных значений частоты и амплитуды становится возможным применение методов линейной теории управления без каких-либо ограничений. Так как предметом анализа обычно являются автоколебательные режимы, имеющие определенное значение частоты и амплитуды, то очевидна большая эффективность метода Е. П. Попова.

Для иллюстрации метода рассмотрим нелинейную систему, состоящую из линейной части и нелинейности, под знаком

которой кроме входной величины содержится ее первая производная:

$$X_2 = f(X_1, \dot{X}_1).$$

Опираясь на метод гармонического баланса, примем $\dot{X}_1 = A \sin \omega t$, следовательно,

$$X_2 = f_{н.э}(A \sin \omega t, A \omega \cos \omega t). \quad (3.18)$$

Введя новую переменную $u = \omega t$ и разложив (3.18) в ряд Фурье, получим, без учета постоянной составляющей и высших гармоник,

$$X_2 \approx \left[\frac{1}{\pi} \int_0^{2\pi} f_{н.э}(A \sin u, \omega A \cos u) \sin u du \right] \sin u + \\ + \left[\frac{1}{\pi} \int_0^{2\pi} f_{н.э}(A \sin u, \omega A \cos u) \cos u du \right] \cos u.$$

Учитывая, что $\sin u = \sin \omega t = X_1/A$; $\cos u = \cos \omega t = \dot{X}_1/(A\omega)$, имеем:

$$X_2 \approx F(A, \omega) X_1 + \frac{G(A, \omega)}{\omega} \dot{X}_1, \quad (3.19)$$

где
$$F(A, \omega) = \frac{1}{\pi A} \int_0^{2\pi} f_{н.э}(A \sin u, A \omega \cos u) \sin u du;$$

$$G(A, \omega) = \frac{1}{\pi A} \int_0^{2\pi} f_{н.э}(A \sin u, A \omega \cos u) \cos u du.$$

При постоянных значениях ω и A выражение (3.19) является линейным. В этом случае нелинейный элемент так же, как линейную часть, можно охарактеризовать передаточной функцией

$$W_{н.э}(p) = K_{н.э}(p)/(H_{н.э}(p)).$$

Передаточная функция всей нелинейной системы

$$W(p) = W_{л}(p) W_{н.э}(p) = \frac{K_{л}(p) K_{н.э}(p)}{H_{л}(p) K_{н.э}(p)}, \quad (3.20)$$

где $W_{л}(p) = K_{л}(p)/H_{л}(p)$ — передаточная функция линейной части системы.

Уравнение (3.20) можно исследовать любым методом, рассмотренным в предыдущей главе применительно к линейным динамическим системам. Из (3.20) следует, что характеристическое уравнение нелинейной системы при фиксированных значениях A и ω имеет вид

$$K_{л}(p) K_{н.э}(p) + H_{л}(p) H_{н.э}(p) = 0. \quad (3.21)$$

Очевидно, что автоколебания возникают всегда, когда уравнение (3.21) содержит два чисто мнимых корня.

В практике исследования нелинейных динамических систем

широко используются также различные графические и графоаналитические методы. Графические методы расчета переходных процессов в нелинейных системах обычно представляют собой ту или иную интерпретацию графического решения нелинейных уравнений в конечных разностях. Точность графических методов зависит от выбранного шага интегрирования. Графические методы исследования динамических систем обладают большой наглядностью, но неудобны для анализа влияния тех или иных факторов на результаты расчетов переходных процессов.

Глава 4

ДИСКРЕТНЫЕ СИСТЕМЫ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

4.1. ОСОБЕННОСТИ ДИСКРЕТНОГО УПРАВЛЕНИЯ. КВАНТОВАНИЕ НЕПРЕРЫВНЫХ ВЕЛИЧИН. МОДУЛЯЦИЯ ИМПУЛЬСОВ

Дискретными называются системы, в которых переменные, характеризующие систему, носят прерывистый дискретный характер. Работа дискретных систем (ДС) связана с воздействием, передачей и преобразованием последовательности импульсов. В отдельные точки ДС сигналы управления поступают в некоторые заданные или случайные промежутки времени. Характерной чертой любой ДС является наличие импульсного элемента (ИЭ), с помощью которого осуществляется преобразование непрерывных величин в последовательность дискретных сигналов — импульсов.

Дискретные системы находят весьма широкое применение в практике управления разнообразными техническими устройствами. Область применения ДС — управление различными электро-механическими и электромагнитными устройствами, электрическими цепями, системами телеизмерения и телеуправления, различными многоканальными системами связи, системами радиоуправления и т. д.

Дискретные системы имеют ряд существенных преимуществ перед непрерывными системами. К важнейшим из них относятся повышенные помехозащищенность и точность системы управления, вытекающие из того факта, что сигнал управления в ДС подвергается помехе только в дискретные моменты времени. В ДС можно более эффективно использовать каналы передачи сигналов управления.

Так же, как и непрерывные, дискретные системы могут быть разомкнутыми и замкнутыми.

Примерами простейших ДС являются электродвигатели и электромагнитные механизмы с системой управления, обеспечивающей импульсный режим работы, устройства телеизмерения, многоканальные системы связи, импульсные счетно-решающие устройства и пр. Примерами более сложных дискретных систем являются

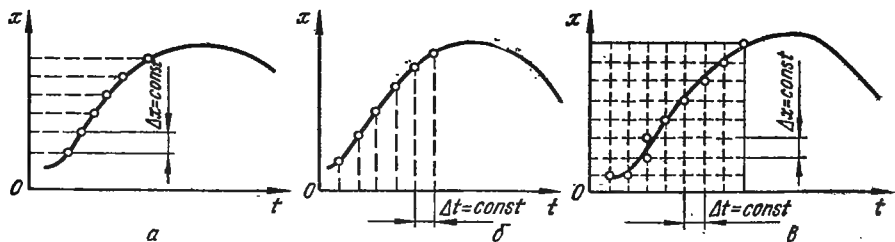


Рис. 4.1. Квантование непрерывных сигналов:
 а — по уровню; б — по времени; в — по уровню и времени

различные системы прерывистого регулирования, импульсные следящие системы, системы автоматического управления, содержащие цифровые вычислительные устройства, системы радиотелеуправления.

Современная теория управления располагает универсальным методом исследования дискретных систем, единым для разнообразных типов ДС. Этот метод основан на использовании специального математического аппарата — дискретного преобразователя Лапласа, который позволил максимально приблизить методологию исследования ДС к методологии исследования непрерывных систем. Использование дискретного преобразования Лапласа дало возможность ввести для дискретных систем ряд понятий, формально аналогичных понятиям теории непрерывных систем, таких как передаточные функции, частотные характеристики и пр. Однако эта аналогия в значительной степени носит формальный характер, вследствие чего теория управления дискретными системами имеет существенные особенности, обусловленные наличием в этих системах импульсных элементов.

Работа ДС связана с *квантованием непрерывных сигналов* — преобразованием непрерывного сигнала, при котором осуществляется дискретизация сигнала по уровню или по времени, или одновременно по уровню и времени (рис. 4.1).

При квантовании по уровню непрерывный сигнал $x(t)$ преобразуется в последовательность дискретных сигналов, фиксированных в произвольные моменты времени (рис. 4.1, а) при условии $\Delta x = \text{const}$. Дискретные системы, в которых реализуются сигналы, квантованные по уровню, называются *релейными системами*. В этих системах квантование по уровню осуществляется релейным элементом, на выходе которого может фиксироваться конечное число уровней (часто 2—3 уровня). Квантование по уровню является нелинейным преобразованием входного сигнала $x(t)$, следовательно, релейные системы относятся к классу нелинейных систем.

При квантовании по времени (рис. 4.1, б) сигналы фиксируются в дискретные моменты времени $\Delta t = \text{const}$. При этом уровни сигнала могут принимать произвольные значения. Дискретные системы, реализующие сигналы, квантованные по времени, называются *импульсными системами* (ИС). В этих системах квантова-

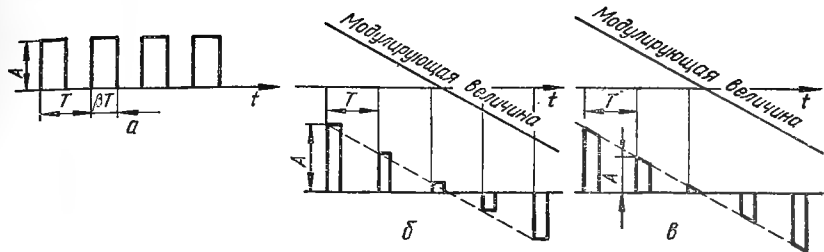


Рис. 4.2. Модуляция импульсов:

a — немодулированная последовательность импульсов; *б* — амплитудно-импульсная модуляция первого рода; *в* — импульсная модуляция второго рода

ние по времени осуществляется импульсным элементом, который в частном случае пропускает входной сигнал $x(t)$ лишь в течение некоторого времени.

При квантовании по уровню и по времени (рис. 4.1, *в*) непрерывный сигнал заменяется дискретными уровнями, ближайшими к значениям непрерывного сигнала в дискретные моменты времени $\Delta t = \text{const}$. Дискретные системы, реализующие сигналы, квантованные по уровню и по времени, называются *релейно-импульсными*, или *цифровыми*. В этих системах квантование по уровню и по времени осуществляется кодоимпульсным модулятором или цифровым вычислительным устройством.

В дальнейшем рассмотрение дискретных систем ограничено импульсными системами. Теория ИС изучает динамические свойства, методы расчета, исследования и построения импульсных систем.

Последовательность импульсов в ИС подвергается импульсной модуляции. Процесс импульсной модуляции состоит в изменении по определенному временному закону какого-либо параметра периодически повторяющихся импульсов. Применительно к немодулированной последовательности импульсов (рис. 4.2, *a*) такими параметрами являются амплитуда импульса A , длительность, или ширина, импульса βT , расстояние между импульсами, или период повторения, T . Величина, определяющая закон модуляции, называется *модулирующей величиной*.

Если по закону изменения модулирующей величины изменяется амплитуда импульсов, то модуляция называется амплитудно-импульсной (АИМ), если изменяется ширина — широтно-импульсной (ШИМ), при изменении периода — временно-импульсной модуляцией (ВИМ).

Вид модуляции, при которой параметры последовательности импульсов изменяются в зависимости от значений модулирующей величины в фиксированные равноотстоящие друг от друга моменты времени, называется *импульсной модуляцией первого рода*. В этом случае модулируемый параметр — амплитуда, ширина или частота импульса — определяется значением модулирующей величины в равноотстоящие дискретные моменты времени. На рис. 4.2, *б* показана амплитудно-импульсная модуляция первого рода.

Вид модуляции, при которой модулируемые параметры последовательности импульсов изменяются в соответствии с текущим значением модулирующей величины, называется *импульсной модуляцией второго рода*. В этом случае модулируемый параметр изменяется в течение времени существования импульса (рис. 4.2, в).

Широкий класс импульсных систем можно представить как совокупность непрерывной части и импульсного элемента. Непрерывная часть представляет динамическую систему той или иной физической природы — электрическую, механическую, электромеханическую, пневматическую и т. д.

Импульсные системы могут быть линейными и нелинейными. В линейных ИС соблюдается принцип суперпозиции: реакция ИС на сумму воздействий равна сумме реакций на каждое воздействие в отдельности. В этих системах параметры импульсного элемента и непрерывной части не зависят от внешних воздействий и переменных, характеризующих состояние системы. К линейным ИС относятся, например, амплитудно-импульсные системы с линейной непрерывной частью и с линейной характеристикой импульсного элемента, под которой понимается зависимость модулируемого параметра выходной последовательности импульсов от соответствующих дискретных значений входной величины. В дальнейшем будут рассматриваться линейные импульсные системы, в которых ИЭ может быть включен до непрерывной части, после нее или между отдельными частями непрерывной системы. В замкнутых ИС импульсный элемент может находиться в прямой части системы, в цепи обратной связи или вне замкнутого контура.

4.2. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ СИСТЕМ ДИСКРЕТНОГО УПРАВЛЕНИЯ

Математический аппарат описания непрерывных систем не может быть использован для описания дискретных систем, так как дискретно представляемые сигналы должны описываться функциями дискретной переменной. В связи с этим для описания дискретных систем в теории управления используются решетчатые функции и разностные уравнения. В теории дискретных систем решетчатые функции являются аналогами непрерывных функций, описывающих непрерывные системы, а разностные уравнения являются аналогами дифференциальных уравнений.

Решетчатой функцией называется функция, получающаяся в результате замены непрерывной переменной дискретной независимой переменной и определенная в дискретные моменты времени nT , $(n+1)T$ (рис. 4.3, а). Непрерывной функции $x(t)$ соответствует решетчатая функция $x[nT]$. Очевидно, что непрерывная функция является огибающей по отношению к решетчатой функции. При заданном значении периода квантования T конкретной непрерывной функции $x(t)$ соответствует одна, вполне определенная, решетчатая функция $x[nT]$. Однако обратного однозначного соответствия между решетчатой и ее непрерывной функцией не

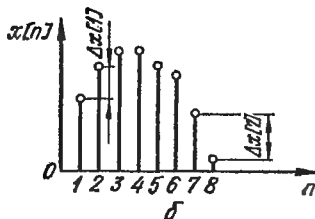
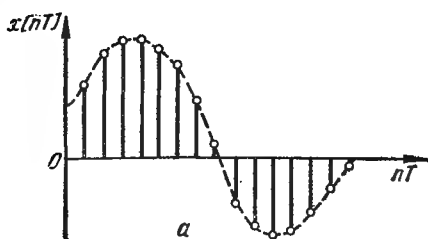


Рис. 4.3. Решетчатые функции:

а — преобразование непрерывной функции в решетчатую; *б* — разность первого порядка

существует, так как через ординаты решетчатой функции можно провести множество огибающих.

Для определения значения решетчатой функции между точками квантования используется понятие *смещенной решетчатой функции* $x[nt, \Delta t]$, где $0 < \Delta t < T$.

Аргумент решетчатой функции целесообразно представить так, чтобы отсчет по шкале времени вести в целочисленных единицах периода квантования T . С этой целью вместо переменной t в аргументе непрерывной функции введем новую переменную $\tau = t/T$. В таком случае непрерывной функции $x(\tau)$ будет соответствовать решетчатая функция $x[n]$.

Связь между значениями решетчатой функции при разных значениях аргумента определяется с помощью конечных разностей, которые являются аналогами производных в дифференциальных уравнениях. *Разностью первого порядка*, или *первой разностью*, называется разность между последующим дискретным значением решетчатой функции и ее предыдущим значением (рис. 4.3, б):

$$\Delta x[n] = x[n+1] - x[n]. \quad (4.1)$$

Разность первого порядка характеризует скорость изменения решетчатой функции и, следовательно, является аналогом первой производной непрерывной функции.

Разность первого порядка, как видно из (4.1), определяется, если известны два последовательных значения решетчатых функций.

Разность второго порядка определяется как разность двух соседних разностей первого порядка:

$$\begin{aligned} \Delta^2 x[n] &= \Delta x[n+1] - \Delta x[n] = \{x[n+2] - x[n+1]\} - \\ &- \{x[n+1] - x[n]\} = x[n+2] - 2x[n+1] + x[n]. \end{aligned}$$

Таким образом, разность второго порядка определяется, если известны три последовательных значения решетчатой функции.

По аналогии с выражением разности второго порядка разность любого m -го порядка будет

$$\Delta^m x[n] = \Delta^{m-1} x[n+1] - \Delta^{m-1} x[n]. \quad (4.2)$$

Раскрывая в (4.2) разности через решетчатые функции, получим

$$\Delta^m x[n] = \sum_{j=0}^m (-1)^j \frac{m!}{j!(m-j)!} x[n+m-j].$$

Математическое описание линейных импульсных систем приводится к виду

$$a_m \Delta^m x[n] + a_{m-1} \Delta^{m-1} x[n] + \dots + a_0 x[n] = 0, \quad (4.3)$$

где a_m — постоянные коэффициенты ($m=0, 1, 2, \dots$). Уравнение (4.3) является линейным разностным уравнением с постоянными коэффициентами — аналог однородного линейного дифференциального уравнения при описании непрерывных динамических систем. Решение (4.3) дает значение дискретной переменной $x[n]$ для каждого периода квантования.

Используя связь значений разностей с дискретными значениями решетчатой функции, уравнение (4.3) можно записать в виде

$$c_m x[n+m] + c_{m-1} x[n+m-1] + \dots + c_0 x[n] = 0 \quad (4.4)$$

или

$$\sum_{i=0}^m c_i x[n+i] = 0. \quad (4.5)$$

Уравнения (4.4) или (4.5) позволяют при известных значениях $x[0], x[1], \dots, x[n]$ последовательно определить значения $x[0+m], x[1+m], \dots, x[n+m]$.

Подобно тому, как в теории непрерывных линейных систем для решения дифференциальных уравнений использовалось преобразование Лапласа, в теории импульсных систем для решения разностных уравнений используется дискретное преобразование Лапласа и его модификация — дискретное z -преобразование.

Обратимся к преобразованию Лапласа для непрерывной функции $x(t)$:

$$x(p) = \int_0^{\infty} x(t) e^{-pt} dt. \quad (4.6)$$

Перейдем в (4.6) от непрерывной функции $x(t)$ к дискретной $x[nt]$, приняв $t=nT$. При этом интеграл превратится в сумму, а сумма приращений времени отобразится периодом квантования T . Тогда

$$x(p) = T \sum_{n=0}^{\infty} x[nt] e^{-pTn}. \quad (4.7)$$

Введя обозначение $e^{pT} = z$, получим

$$x(z) = T \sum_{n=0}^{\infty} x[nT] z^{-n}. \quad (4.8)$$

Это уравнение представляет собой дискретное преобразование Лапласа, в котором выражение

$$x(z) = \sum_{n=0}^{\infty} x[nT] z^{-n} \quad (4.9)$$

называется z -преобразованием. Оно лежит в основе метода решения разностных уравнений. Как видно из (4.8) и (4.9), дискретное преобразование Лапласа $X(z)$ отличается от z -преобразования на наличием множителя T .

Подобно тому, как при анализе непрерывных систем преобразование Лапласа позволяет перейти от дифференциальных уравнений к алгебраическим, более простым, так при анализе дискретных систем z -преобразование позволяет перейти от разностных уравнений к алгебраическим и существенно упростить анализ динамики дискретных систем.

В (4.9) функция $x[nt]$ называется *оригиналом решетчатой функции*, а $X(z)$ — *изображением*. Для обратного перехода от изображения к оригиналу, т. е. для нахождения исходной решетчатой функции по ее изображению используется обратное z -преобразование:

$$x[nT] = \frac{1}{2\pi j} \oint X(z) z^{n-1} dz. \quad (4.10)$$

Для нахождения z -изображения в любые другие моменты времени, отличные от $t=nT$, используется прямое модифицированное z -преобразование

$$X(z, \varepsilon) = \sum_{n=0}^{\infty} x[(n + \varepsilon)T] z^{-n},$$

где ε — смещение по отношению к моменту квантования.

Обратное модифицированное z -преобразование имеет вид

$$x[(n + \varepsilon)T] = \frac{1}{2\pi j} \oint X(z, \varepsilon) z^{n-1} dz.$$

Рассмотрим порядок использования z -преобразования для нахождения решений линейных разностных уравнений. Пусть имеем неоднородное линейное разностное уравнение, приведенное к виду, аналогичному (4.5):

$$\sum_{i=0}^N a_i y[n+i] = \sum_{i=0}^M b_i x[n+i], \quad (4.11)$$

где a_i, b_i — постоянные коэффициенты; $y[n+i]$ — искомая выходная дискретная величина; $x[n+i]$ — входная дискретная переменная; n, i — целые положительные числа.

Дополнив обе части уравнения (4.11) множителем $z^{-(n+i)} = z^{-n} z^{-i}$, получим

$$Y(z) \sum_{i=0}^N a_i z^{-i} = X(z) \sum_{i=0}^M b_i z^{-i},$$

откуда получаем

$$Y(z) = X(z) \frac{\sum_{i=0}^M b_i z^{-i}}{\sum_{i=0}^N a_i z^{-i}}.$$

Далее, с помощью обратного z -преобразования находим выходную дискретную переменную

$$y[n+i] = \frac{1}{2\pi j} \oint Y(z) z^{n-1} dz.$$

Необходимо подчеркнуть, что все рассуждения справедливы при нулевых начальных условиях, когда в начальный момент времени решетчатая функция равна нулю.

4.3. УРАВНЕНИЕ И ПЕРЕДАТОЧНЫЕ ФУНКЦИИ ИМПУЛЬСНЫХ СИСТЕМ

Анализ импульсной системы ограничим случаем, когда ее можно представить как совокупность импульсного элемента ИЭ и линейной непрерывной части НЧ. Рассмотрим сначала разомкнутую импульсную систему (рис. 4.4, а, б).

Импульсный элемент ИЭ преобразует непрерывную функцию $x(t)$ в решетчатую $x[lT]$, которая поступает на вход непрерывной части НЧ.

В силу инерционности НЧ на ее выходе будут генерироваться непрерывные сигналы $y(t)$. Чтобы использовать математический аппарат описания дискретных систем, введем на выходе НЧ фиктивный импульсный элемент ИЭ1, работающий синхронно с ИЭ на входе системы. z -Изображения входной $X(z)$ и выходной $Y(z)$ величин имеют вид

$$X(z) = \sum_{l=0}^{\infty} x[lT]z^{-l}; \quad Y(z) = \sum_{n=0}^{\infty} y[nT]z^{-n}. \quad (4.12)$$

Так как непрерывная часть системы линейна, то непрерывная функция $y(t)$ может быть определена как переходный процесс в системе, находящейся под воздействием последовательности импульсов $x[lT]$. Как было отмечено в гл. 2, переходный процесс в линейной динамической системе в общем случае может быть выражен через импульсную переходную функцию $\omega(t)$ с использованием интеграла свертки (2.87). Применительно к рассматриваемой

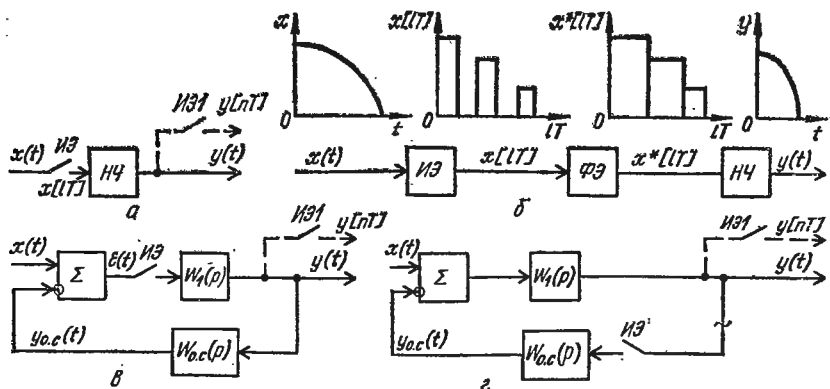


Рис. 4.4. Импульсные системы:

а — разомкнутая без фиксирующего элемента; б — разомкнутая с фиксирующим элементом нулевого порядка; в — замкнутая с импульсным элементом; г — замкнутая с импульсным элементом в цепи обратной связи

тому случаю в интеграле (2.87) имеем $x(t) = y(t)$; $\tau = lT$; $u(\tau) = x[lT]$. Таким образом, получаем

$$y(t) = \sum_{l=0}^{\infty} \omega(t - lT) x[lT]. \quad (4.13)$$

На выходе импульсного элемента ИЭ1

$$y[nT] = \sum_{l=0}^{\infty} \omega[nT - lT] x[lT]. \quad (4.14)$$

С учетом (4.12) и (4.14) получим выражение для z -изображения выходной величины:

$$Y(z) = \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} \omega[nT - lT] x[lT] z^{-n}. \quad (4.15)$$

Введем обозначение $m = n - l$, тогда $z^{-n} = z^{-m} z^{-l}$. С учетом этого из (4.15) вытекает

$$Y(z) = W(z) X(z), \quad (4.16)$$

где

$$W(z) = \sum_{m=0}^{\infty} \omega[mT] z^{-m}. \quad (4.17)$$

Выражение (4.17) представляет собой z -изображение импульсной переходной функции $\omega(t)$, или функции веса. Уравнение (4.16) устанавливает связь z -изображений выходной и входной величин

$$W(z) = Y(z)/X(z). \quad (4.18)$$

По аналогии с непрерывными системами отношение z -преобразованной выходной величины звена (системы) к z -преобразованной входной величине при нулевых начальных условиях называется *дискретной передаточной функцией звена (системы)*.

Из приведенных рассуждений видно, что для определения дискретной передаточной функции системы необходимо, используя передаточную функцию непрерывной части системы, определить импульсную переходную функцию, или функцию веса $\omega(t)$, перейти к весовой последовательности импульсов $\omega[mT]$ и просуммировать ряд (4.17). Из (4.16) видно, что, зная z -изображение входной величины и дискретную передаточную функцию, можно найти z -изображение выходной величины импульсной системы. Таким образом, получаем аппарат исследования ИС — дискретные передаточные функции, — аналогичный аппарату передаточных функций непрерывных линейных систем. Следует, однако, подчеркнуть, что эта аналогия носит формальный характер.

Импульсная система может иметь в своем составе фиксирующий элемент ФЭ (рис. 4.4, б), который сохраняет амплитуду каждого предшествующего импульса до появления следующего импульса. Фиксирующий элемент такого типа называется *фиксирующим элементом нулевого порядка* и может быть физически реализован, например, с помощью конденсатора большой емкости.

Характер сигналов в различных точках системы в этом случае показан на рис. 4.4, б, из которого видно, что на выходе ФЭ генерируется ступенчатый сигнал с амплитудой, изменяющейся в моменты IT .

Дискретная передаточная функция импульсной системы с фиксирующим элементом нулевого порядка имеет вид

$$W(z) = \frac{z-1}{z} Z\left\{\frac{W(p)}{p}\right\}, \quad (4.19)$$

где $z=e^{Tp}$ — комплексная переменная; Z — символ z -преобразования; $W(p)$ — передаточная функция непрерывной части импульсной системы. Следует подчеркнуть, что под знак z -преобразования должна попасть передаточная функция всей непрерывной части системы. Такая оговорка необходима для импульсных систем, в которых непрерывная часть представлена несколькими последовательно соединенными структурными звеньями.

Передаточную функцию замкнутой импульсной системы рассмотрим сначала для случая, когда импульсный элемент включен в прямой цепи на входе непрерывной части с передаточной функцией $W_1(p)$ (рис. 4.4, в). Так же, как и в разомкнутой системе, несмотря на наличие импульсного элемента ИЭ, на выходе непрерывной части $W_1(p)$ рассматриваемой импульсной системы вследствие инерционности НЧ будет непрерывный сигнал $y(t)$, который через обратную связь с передаточной функцией $W_{o.c}(p)$ поступает на вход системы. Выделив в непрерывной функции $y(t)$ решетчатую функцию $y[nT]$ и выполнив z -преобразования, найдем дискретное значение сигнала ошибки на входе ИЭ:

$$\varepsilon(z) = X(z) - Y_{o.c}(z). \quad (4.20)$$

С другой стороны, величина $\varepsilon(z)$ может быть определена из выражения

$$Y(z)/\varepsilon(z) = W_1(z), \text{ т. е. } \varepsilon(z) = Y(z)/W_1(z). \quad (4.21)$$

При отсутствии фиксирующего устройства z -изображение величины на выходе обратной связи имеет вид

$$Y_{o.c}(z) = Z\{W_1(p)W_{o.c}(p)\}\varepsilon(z). \quad (4.22)$$

С учетом выражений (4.21) и (4.22) уравнение (4.20) принимает вид

$$\frac{Y(z)}{W_1(z)} = X(z) - Z\{W_1(p)W_{o.c}(p)\} \frac{Y(z)}{W_1(z)},$$

откуда получаем выражение передаточной функции рассматриваемой замкнутой импульсной системы без фиксирующего устройства:

$$W(z) = \frac{Y(z)}{X(z)} = \frac{W_1(z)}{1 + Z\{W_1(p)W_{o.c}(p)\}}. \quad (4.23)$$

Для формальной аналогии (4.23) с выражением передаточной функции замкнутой линейной непрерывной системы представим последнее выражение в виде

$$W(z) = \frac{W_1(z)}{1 + W_1 W_{o.c}(z)},$$

где $W_1 W_{o.c}(z) \equiv Z\{W_1(p) W_{o.c}(p)\}$.

Если на входе непрерывной части импульсной системы будет находиться фиксирующее устройство нулевого порядка, то, проведя аналогичные рассуждения, найдем

$$W(z) = \frac{Y(z)}{X(z)} = \frac{\frac{z-1}{z} Z \left[\frac{W_1(p)}{p} \right]}{1 + \frac{z-1}{z} Z \left[\frac{W_1(p) W_{o.c}(p)}{p} \right]}.$$

Выражение передаточной функции замкнутой импульсной системы существенно зависит от того, где включен импульсный элемент внутри контура системы. Пусть, например, он включен перед непрерывной частью обратной связи с передаточной функцией $W(p)$ (рис. 4.4, з). Разомкнув мысленно обратную связь на входе импульсного элемента, можно записать выражение для z -изображения выходной величины:

$$Y(z) = Z\{W_1(p) X(p)\} - Z\{W_{o.c}(p) W_1(p)\} Y(z),$$

откуда

$$Y(z) = \frac{W_1 X(z)}{1 + Z\{W_{o.c}(p) W_1(p)\}},$$

где $W_1 X(z) = Z\{W_1(p) X(p)\}$.

4.4. УСТОЙЧИВОСТЬ ИМПУЛЬСНЫХ СИСТЕМ

Так же, как и в непрерывных системах, под устойчивостью импульсных систем понимается их способность с течением времени возвращаться в равновесное состояние после того, как это состояние было нарушено возмущающими воздействиями. Дискретный характер сигналов управления, передаваемых в импульсных системах, существенно усложняет анализ ИС на устойчивость, хотя рассмотренный выше аппарат передаточных функций дискретных систем делает порядок анализа ИС формально аналогичным анализу на устойчивость линейных непрерывных систем.

Для суждения об устойчивости импульсной системы необходимо проанализировать характер вынужденного движения в системе.

Общее уравнение движения импульсной системы, находящейся под влиянием возмущения, можно привести к виду

$$\sum_{i=0}^m c_i x[(n+i)T] = \sum_{i=0}^h b_i y[(n+i)T]. \quad (4.24)$$

Левая часть этого уравнения вытекает из (4.5) при замене целочисленного аргумента $n+i$ на аргумент $(n+i)T$, а правая часть отражает эффект возмущающего воздействия $y(t)$.

В развернутой форме (4.24) будет иметь вид

$$c_0x[nT] + c_1x[(n+1)T] + \dots + c_mx[(n+m)T] = \\ = b_0y[nT] + b_1y[(n+1)T] + \dots + b_ky[(n+k)T]. \quad (4.25)$$

Общее решение уравнения (4.25)

$$x[nT] = x_B[nT] + x_0[nT],$$

где $x_B[nT]$ — вынужденное движение системы; $x_0[nT]$ — свободное движение системы.

Об устойчивости системы следует судить по сходимости или не-сходимости составляющей $x_0[nT]$, определяемой из решения дискретного однородного уравнения

$$c_0x[nT] + c_1x[(n+1)T] + \dots + c_mx[(n+m)T] = 0. \quad (4.26)$$

Решение его будем искать в виде

$$x_0[nT] = u^{nT}, \quad u^T = z, \quad \text{т. е. } x_0[nT] = z^n.$$

С учетом этого (4.26) примет вид

$$c_0u^{nT} + c_1u^{(n+1)T} + \dots + c_mu^{(n+m)T} = 0.$$

После сокращения на u^{nT} ($u^{nT} \neq 0$) получим

$$c_0 + c_1u^T + \dots + c_mu^{mT} = 0. \quad (4.27)$$

Обозначив $u^T = z$, получим характеристическое уравнение свободного движения импульсной системы

$$c_0 + c_1z + \dots + c_mz^m = 0. \quad (4.28)$$

Решение уравнения (4.28) имеет вид

$$x_0[nT] = a_1z_1^n + a_2z_2^n + \dots + a_mz_m^n, \quad (4.29)$$

где $z_1, z_2, \dots, z_m$ — корни характеристического уравнения; $a_1, a_2, \dots, a_m$ — постоянные коэффициенты, определяемые из начальных условий $x_0[0T]$; $x_0[1T]$; $x_0[2T]$; ...; $x_0[(m-1)T]$.

Очевидно, что решение (4.29) будет сходящимся, а система, имеющая характеристическое уравнение (4.28), будет устойчивой, если

$$\lim_{n \rightarrow \infty} x_0[nT] = 0. \quad (4.30)$$

Это условие будет выполнено, если каждое слагаемое правой части (4.29) удовлетворяет условию

$$\lim_{n \rightarrow \infty} z_m^n = 0, \quad (4.31)$$

которое выполняется только тогда, когда все корни $z_1, z_2, \dots, z_m$ по модулю меньше единицы, т. е.

$$\lim_{n \rightarrow \infty} z_m^n = 0 \quad \text{при } |z_m| < 1 \quad \text{и} \quad \lim_{n \rightarrow \infty} z_m^n \neq 0 \quad \text{при } |z_m| > 1.$$

Таким образом, необходимым и достаточным условием устойчивости импульсной системы является условие

$$|z_m| < 1 \quad (4.32)$$

для всех корней характеристического уравнения, соответствующего разностному уравнению системы.

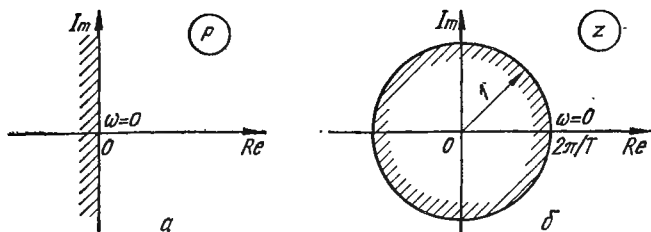


Рис. 4.5. Геометрическая интерпретация условий устойчивости систем: а — линейной непрерывной; б — линейной импульсной

Учитывая, что $z = e^{pT}$, преобразуем плоскость комплексного переменного p в комплексную плоскость корней z (рис. 4.5). Граница устойчивости линейной непрерывной системы (рис. 4.5, а) определяется условием $p = j\omega$. При этом система устойчива, если все корни выходят слева от мнимой оси. Эта граница на плоскости корней z (рис. 4.5, б) отображается в окружность единичного радиуса. Действительно, $z = e^{j\omega T}$ при $p = j\omega$. При $z = 1$ и при изменении частоты в пределах $0 < \omega < 2\pi/T$ переменная z описывает на плоскости z окружность единичного радиуса. При дальнейшем изменении ω вплоть до $2\pi/T$ переменная z всякий раз описывает окружность единичного радиуса при возрастании ω на $2\pi/T$.

Нетрудно заметить, что область, заключенная внутри круга единичного радиуса на плоскости z , соответствует значению параметров импульсной системы, при которых система устойчива, так как левая полуплоскость плоскости p отображается во внутреннюю плоскость круга единичного радиуса плоскости z . Действительно, если $p = -c + j\omega$, то непрерывная система устойчива и с увеличением значения c запас устойчивости увеличивается. Применительно к плоскости z в этом случае $z = e^{pT} = e^{-cT}e^{j\omega T}$ и при $c \rightarrow \infty$ имеем $z \rightarrow 0$, т. е. запас устойчивости импульсной системы возрастает, если корни z перемещаются с окружности единичного радиуса внутрь.

Следовательно, в устойчивой импульсной системе корни на плоскости z располагаются внутри круга единичного радиуса $|z| < 1$. Это условие аналогично условию существования только левых корней для обеспечения устойчивости линейной непрерывной системы.

На основании сформулированного общего условия устойчивости линейных импульсных систем разработан ряд критериев устойчивости ИС, позволяющих вынести суждение об устойчивости по соотношению корней характеристического уравнения без вычисления самих корней. Критерии устойчивости импульсных систем управления формально аналогичны соответствующим критериям устойчивости линейных непрерывных систем. В частности, аналог критерия Рауса — Гурвица позволяет вынести суждение об устойчивости импульсной системы на основе анализа соотношения корней характеристического уравнения системы

$$G(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_m z^m = 0. \quad (4.33)$$

Для определения устойчивости ИС из коэффициентов уравнения (4.33) по определенному правилу составляются определители, которые в устойчивой ИС должны отвечать определенным требованиям. Вид этих определителей более сложен, чем определителей Гурвица для непрерывных линейных систем, так как в их состав входят не только сами коэффициенты, но и их сопряженные значения, т. е. коэффициенты, расположенные обратнoсимметрично по отношению к коэффициентам уравнения (4.33).

Анализ устойчивости импульсных систем по аналогу критерия Рауса — Гурвица можно свести к процедурам, полностью аналогичным процедурам, имевшим место при анализе непрерывных линейных систем. С этой целью используется билинейное преобразование

$$z = \frac{1+w}{1-w} \quad \text{или} \quad w = \frac{z-1}{z+1},$$

которое отображает круг единичного радиуса плоскости z в левую полуплоскость плоскости w . С учетом введенного преобразования получается характеристический полином как однозначная функция w и в отношении этого полинома условия устойчивости полностью совпадают с условиями Гурвица для непрерывных линейных систем.

В качестве примера рассмотрим условие устойчивости импульсной системы второго порядка, имеющей характеристическое уравнение

$$A_0 z^2 + A_1 z + A_2 = 0. \quad (4.34)$$

Выполнив подстановку $z = \frac{1+w}{1-w}$, приведем характеристическое уравнение к виду

$$B_0 w^2 + B_1 w + B_2 = 0, \quad (4.35)$$

где $B_0 = A_0 - A_1 - A_2$; $B_1 = 2(A_0 - A_2)$; $B_2 = A_0 + A_1 + A_2$.

По критерию Гурвица для непрерывных систем линейная система, имеющая характеристическое уравнение (4.35), устойчива, если все коэффициенты этого уравнения положительны, т. е. если $B_0 > 0$; $B_1 > 0$; $B_2 > 0$.

Учитывая полученные соотношения коэффициентов B и A , находим условие устойчивости импульсной линейной системы, имеющей характеристическое уравнение (4.34):

$$A_0 - A_1 - A_2 > 0; \quad 2(A_0 - A_2) > 0; \quad A_0 + A_1 + A_2 > 0.$$

Аналог критерия устойчивости Михайлова импульсных систем базируется на анализе значения результирующего угла поворота на плоскости z вектора, соответствующего характеристическому уравнению (4.33) $G(z) = 0$. Это уравнение имеет m корней $z_1, z_2, \dots, z_m$. При известных корнях (4.33) можно представить в виде произведения сомножителей

$$G(z) = a_m (z - z_1) (z - z_2) \dots (z - z_m) = 0. \quad (4.36)$$

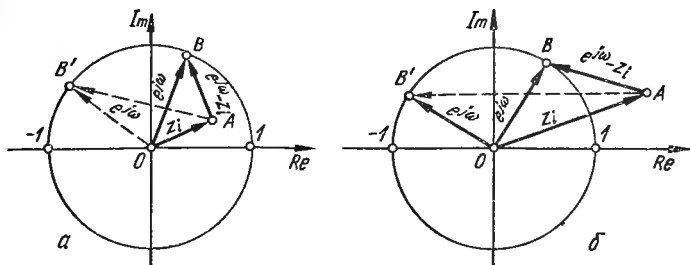


Рис. 4.6. Критерий устойчивости импульсных систем — аналог критерия Михайлова

При $z = e^{j\omega}$

$$G(e^{j\omega}) = a_m(e^{j\omega} - z_1)(e^{j\omega} - z_2) \dots (e^{j\omega} - z_m) = 0. \quad (4.37)$$

При изменении ω в диапазоне $-\pi < \omega < \pi$ суммарный угол поворота вектора $G(z)$ равен сумме углов поворота отдельных векторов — сомножителей $(e^{j\omega} - z_i)$, $i = \overline{1, m}$.

В случае устойчивой импульсной системы все корни z_i , $i = \overline{1, m}$, располагаются внутри круга единичного радиуса, который описывается концом вектора $e^{j\omega}$ при изменении частоты в диапазоне $-\pi < \omega < \pi$. В случае неустойчивой импульсной системы все или часть корней расположатся за пределами круга единичного радиуса (рис. 4.6).

При $|z_i| < 1$, т. е. в случае устойчивой ИС, конец вектора $\overline{OA} = z_i$ будет находиться внутри круга единичного радиуса, а конец вектора $\overline{OB} = e^{j\omega}$ при изменении частоты в пределах $-\pi < \omega < \pi$ скользит по этой окружности (рис. 4.6, а). Угол поворота вектора $\overline{AB} = e^{j\omega} - z_i$, соответствующего i -му сомножителю ($i = \overline{1, m}$) выражения (4.37), при изменении частоты в указанном диапазоне составит 2π . Следовательно, суммарный угол поворота вектора $G(z) = G(e^{j\omega})$, обусловленный всеми сомножителями произведения (4.37), при изменении частоты в пределах $-\pi < \omega < \pi$ составит $2\pi m$.

При $|z_i| > 1$, т. е. если система неустойчива, конец вектора $\overline{OA} = z_i$ окажется за пределами круга единичного радиуса (рис. 4.6, б). В этом случае вектор $\overline{AB} = e^{j\omega} - z_i$ при изменении частоты $-\pi < \omega < \pi$ совершит одно колебательное движение, причем его поворот в положительном направлении (против часовой стрелки) будет равен повороту в отрицательном направлении (по часовой стрелке), а результирующий угол поворота вектора $\overline{AB}$ будет равен нулю. Таким образом, каждый из корней $|z_i| > 1$, расположенный за пределами круга единичного радиуса, уменьшит результирующий угол поворота вектора $G(z) = G(e^{j\omega})$ на 2π .

На основании изложенного аналог критерия Михайлова формулируется следующим образом: линейная импульсная система будет устойчива, если вектор $G(e^{pT}) = G(e^{j\omega T})$ при изменении час-

тоты ω от $-\pi$ до π (при $T=1$), вращаясь против часовой стрелки, совершает результирующий поворот на угол $2\pi m$, где m — порядок полинома характеристического уравнения импульсной системы.

Критерий устойчивости импульсных систем — аналог критерия Найквиста — основан на связи между формой дискретной передаточной функции разомкнутой системы $W(z)$ и свойством устойчивости замкнутой системы. Идея критерия и его доказательство аналогичны критерию Найквиста, рассмотренному в гл. 2 для линейных непрерывных систем. Аналог критерия Найквиста формулируется так: замкнутая импульсная система, устойчивая в разомкнутом состоянии, будет устойчивой, если годограф вектора $\bar{W}(z)$, соответствующий дискретной передаточной функции разомкнутой системы при изменении частоты от 0 до $2\pi/T$, не охватывает точку на плоскости z с координатами $(-1; j0)$.

4.5. КАЧЕСТВО УПРАВЛЕНИЯ В ИМПУЛЬСНЫХ СИСТЕМАХ

Проблема качества управления в дискретных, в частности в импульсных, системах аналогична проблеме качества управления в непрерывных системах. Под качеством управления ИС понимается форма переходного процесса, которая при качественном управлении должна отвечать заданным требованиям — устойчивости, быстродействию, заданной форме дискретной выходной величины. В связи с этим после анализа ИС на устойчивость необходимо стабилизировать систему, если она неустойчива, и построить график переходного процесса, если система устойчива. Если в устойчивости ИС форма переходного процесса не отвечает заданным требованиям, то необходимо осуществить коррекцию ИС, сущность которой аналогична целям, задачам и методам коррекции непрерывных линейных динамических систем.

Построение переходных процессов в дискретных системах с использованием решетчатых функций и z -преобразования рассмотрено в п. 4.2. Другой метод построения переходных процессов в замкнутых импульсных системах базируется на использовании частотной характеристики $W(j\omega)$ замкнутой ИС. Характеристика $W(j\omega)$ получается из дискретной передаточной функции $W(z)$ замкнутой ИС путем замены $z = e^{j\omega}$.

Построение переходных процессов в импульсной системе с использованием частотной характеристики $W(j\omega)$ аналогично построению переходных процессов в линейных непрерывных системах по методу трапецевидных вещественных частотных характеристик, рассмотренному в п. 2.5. Однако наличие импульсного элемента вносит некоторые особенности. Приведем порядок определения дискретной выходной величины $y[n]$ в импульсной системе управления по методу, основанному на использо-

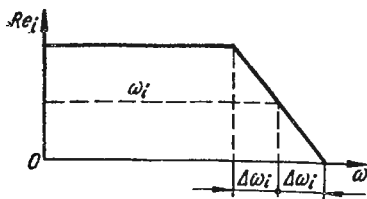


Рис. 4.7. Типовая вещественная частотная характеристика замкнутой импульсной системы

вании вещественной частотной характеристики замкнутой ИС. Отделив вещественную часть $\text{Re}W(j\omega)$ от мнимой, построим зависимость вещественной частотной характеристики от частоты ω и разобьем ее на отдельные типовые трапеции так же, как это делалось в непрерывных системах. Параметры типовой трапециевидной вещественной частотной характеристики $\text{Re}_i(\omega)$ показаны на рис. 4.7. Не останавливаясь на доказательстве, приведем выражение для определения выходной дискретной величины на базе параметров характеристики $\text{Re}(\omega)$:

$$y[n] = \sum_{k=0}^{n-l-m} c_k x[n - (l - m + k)],$$

где $x[n]$ — входная величина ИС; l — порядок полинома знаменателя передаточной функции замкнутой ИС; m — порядок полинома числителя передаточной функции замкнутой ИС; n — дискретные значения аргумента входной и выходной величин.

Коэффициенты c_k определяются на основе параметров типовых характеристик $\text{Re}(\omega)$ по уравнению

$$c_k = \sum_{i=1}^j S_i \frac{\sin \omega_i k}{\omega_i k} \frac{\sin \Delta \omega_i k}{\Delta \omega_i k},$$

где j — число трапеций, на которые разбита характеристика $\text{Re}(\omega)$; i — номер трапеции ($i = \overline{1, j}$), S_i — площадь i -й трапеции.

Для определения значений коэффициентов c_k удобно пользоваться таблицей значений $\sin x/x$.

Коррекция импульсных систем отличается от коррекции непрерывных систем тем, что характеристики ИС зависят не только от структуры и параметров непрерывной части ИС, но и от характеристики импульсного элемента. В связи с этим коррекцию ИС, т. е. изменение характеристик импульсной системы управления с целью обеспечения устойчивости и требуемого качества управления, можно осуществлять как коррекцией характеристик непрерывной части, так и изменением процессов прерывания, т. е. коррекцией характеристик импульсных элементов. Коррекция ИС подразделяется на непрерывную и дискретную.

Наличие процессов прерывания в ИС усложняет определение последовательного корректирующего устройства, вводимого в непрерывную часть ИС. Действительно, пусть дискретная передаточная функция исходной ИС равна $W_{\text{исх}}(z)$. После введения последовательного корректирующего устройства с передаточной функцией $W_{\text{к.у}}(z)$ передаточная функция ИС примет вид $W(z) = (W_{\text{исх}} W_{\text{к.у}})(z)$.

Следует подчеркнуть, что $(W_{\text{к.у}} W_{\text{исх}})(z) \neq W_{\text{к.у}}(z) W_{\text{исх}}(z)$, так как вследствие прерываний в импульсной системе прохождение сигнала управления через отдельные звенья $W_{\text{исх}}(z)$ и $W_{\text{к.у}}(z)$ не эквивалентно прохождению такого сигнала через последовательное соединение $(W_{\text{к.у}} W_{\text{исх}})(z)$. Таким образом, дискретная передаточная функция разомкнутой системы не является произведением дискретных функций, полученных из соответствующих передаточ-

ных функций отдельных непрерывных звеньев, образующих систему.

При коррекции ИС в дискретную цепь вводятся дополнительные цепи, содержащие импульсные элементы, например импульсные фильтры. В этом случае передаточная функция разомкнутой ИС $W(z) = W_{\text{исх}}(z) W_{\text{к.у}}(z)$, и метод коррекции ничем не отличается от метода последовательной коррекции непрерывных систем.

Передаточная функция $W_{\text{к.у}}(z)$ и соответствующее ей реальное физическое звено получаются сравнением частотных характеристик желаемой системы с характеристиками исходной системы.

Глава 5

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

5.1. ОСНОВНЫЕ ПОНЯТИЯ СИСТЕМ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Оптимальным называется управление, осуществляемое наилучшим по определенным показателям образом. Системы, реализующие оптимальное управление, называются *оптимальными*. Организация оптимального управления основана на выявлении и реализации предельных возможностей систем.

Теория оптимального управления возникла в середине 50-х годов текущего столетия на базе задач теории автоматического регулирования. Позже оптимальное управление распространилось на объекты производственно-экономического характера. Значительное развитие теория оптимального управления получила в результате работ советских ученых — академика Л. С. Понтрягина и его сотрудников В. Г. Болтянского, Р. В. Гамкрелидзе, Е. Ф. Мищенко. Эти работы широко используются для синтеза оптимальных систем управления как в нашей стране, так и за рубежом. Значительный вклад в развитие оптимального управления внесли советские ученые А. А. Фельдбаум, Н. И. Красовский, А. М. Летов, В. С. Пугачев и др., а также зарубежные ученые Р. Беллман, Р. Калман, М. Атанс, П. Фалб и др.

Важное направление развития систем оптимального управления — самонастраивающиеся системы, реализующие автоматическую оптимизацию. В этой области широко известны работы советских ученых В. М. Глушкова, П. И. Чинаева, В. М. Кунцевича, американских ученых Дж. Траскела, Л. Брауна и др.

При разработке систем оптимального управления одним из важнейших шагов является формулировка *критерия оптимальности*, под которым понимается основной показатель, определяющий задачу оптимизации. Именно по этому критерию оптимальная система должна функционировать наилучшим образом.

В качестве критериев оптимальности выступают разнообразные технические и технико-экономические показатели, выражающие технико-экономическую выгоду или, наоборот, потери. В первом

случае оптимальное управление должно обеспечивать максимум критерия оптимальности, например производительности, коэффициента полезного действия, прибыли и т. д., при заданных реальных условиях работы и ограничениях. Во втором случае оптимальное управление должно обеспечивать минимум критерия оптимальности при заданных ограничениях, например расхода энергии, топлива, финансовых ресурсов и т. п.

В силу противоречивости требований, предъявляемых к системам автоматического управления, выбор критерия оптимальности обычно превращается в сложную задачу, имеющую неоднозначное решение. Например, оптимизация автоматической системы по критерию надежности может повлечь за собой увеличение стоимости системы, ее усложнение. С другой стороны, упрощение системы снизит ряд других ее показателей. К тому же, не всякое оптимальное решение, синтезированное теоретически, можно реализовать на практике на базе достигаемого уровня техники. В связи с этим чаще всего управление синтезируется оптимальным по какому-либо основному критерию, а остальные, определяющие качество функционирования системы, ограничиваются областью допустимых значений. Это упрощает и делает более определенной задачу поиска оптимальных решений при разработке оптимальных систем. Вместе с тем усложняется задача выбора конкурирующих вариантов систем, так как они сравниваются по различным критериям, а оценка системы не имеет однозначного ответа. Действительно, без тщательного анализа множества противоречивых, зачастую неформализуемых факторов трудно ответить, например, на такой вопрос — какая из систем лучше: более надежная или менее дорогая?

Критерий оптимальности обычно представляется в виде некоторого функционала. Функционал в таком случае можно определить как функцию, аргументы которой связаны с критериями оптимальности и сами являются функциями переменных.

Обобщенный критерий качества работы динамической системы можно представить в виде функционала качества

$$J = F(C, K, H, W, M, V), \quad (5.1)$$

где C — стоимость разработки, создания и эксплуатации системы; K — качество функционирования; H — надежность; W — потребляемая энергия; M — масса; V — объем.

Каждый из перечисленных аргументов функционала J является самостоятельным критерием качества и функцией многих переменных.

Нетрудно видеть, что оптимизация системы по функционалу (5.1) практически невозможна, так как его значение определяется многими противоречивыми факторами. Поэтому в теории управления используются функционалы, характеризующие отдельные показатели качества работы систем управления. Применительно к динамическим системам функционал качества в общем случае включает в себя координаты выхода $X\{x_1, x_2, \dots, x_n\}$, координаты

управляющих воздействий $U\{u_1, u_2, \dots, u_r\}$ и координаты возмущающих действий $F\{f_1, f_2, \dots, f_k\}$, где n, r, k — число соответствующих переменных. При этом функционал качества систем в векторной форме имеет вид

$$J = \int_{t_1}^{t_2} \Psi(X, U, F) dt \quad (5.2)$$

или

$$J = \int_{t_1}^{t_2} \Psi(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; f_1, f_2, \dots, f_k) dt, \quad (5.3)$$

где t_1, t_2 — интервалы времени, в которых определяется функционал.

Оптимальное поведение или состояние системы обеспечивается тогда, когда функционал (5.2) достигает своего экстремума $J = \text{extr}$ — максимума или минимума, в зависимости от физического смысла переменных.

В практике разработки и исследования динамических систем наиболее часто встречаются две задачи: 1) синтез системы, оптимальной по быстродействию; 2) синтез системы, оптимальной по точности. В первом случае необходимо обеспечить минимум времени переходного процесса $x_i(t)$, во втором — минимум среднеквадратичной ошибки (отклонения координаты $x_i(t)$ от заданного значения) при заданных или случайных воздействиях.

Не останавливаясь на глубоком обосновании выбора конкретного критерия оптимального управления, укажем некоторые типы критериев оптимальности, подразделяя их в зависимости от режима работы динамической системы.

Критерий оптимальности системы, оптимальной по быстродействию,

$$J = \int_{t_1}^{t_2} f(t) dt \rightarrow \min. \quad (5.4)$$

При $f(t) = [1]$ имеем $J = t_2 - t_1 = \min$.

Система, оптимальная по точности в динамических режимах, характеризуется следующим функционалом качества:

$$J = \int_0^T \Delta x^2(t) dt \rightarrow \min, \quad (5.5)$$

где $\Delta x(t)$ — отклонение выходной величины от заданного значения; $0, T$ — рассматриваемый интервал времени.

Квадрат переменной под знаком интеграла обеспечивает независимость критерия от знака отклонения выходной переменной. Критерий (5.5) характеризует суммарную ошибку в динамической системе управления за время переходного процесса. Минимизация функционала (5.5) означает ограничение отклонений регулируемой величины и времени их существования.

В ряде случаев используется обобщенная интегральная оценка, ограничивающая отклонение не только выходной переменной, но и ее производных.

Оптимальная по условиям инвариантности (независимости) от возмущения $f_i(t)$ система характеризуется показателем качества

$$J = \int_0^T x_i^2(t + \tau) f_i^2(t) dt \rightarrow \min, \quad (5.6)$$

где $x_i(t)$ — переменная, для которой требуется обеспечить независимость от возмущения $f_i(t)$; τ — приращение времени.

Для многомерных объектов управления, т. е. для объектов с несколькими управляемыми переменными, необходимо обеспечить независимость некоторой переменной $x_i(t)$ от другой переменной $x_k(t)$. Функционал качества в таких системах

$$J = \int_0^T x_i^2(t + \tau) x_k^2(t) dt \rightarrow \min. \quad (5.7)$$

Оптимальные системы, функционирующие по критерию (5.6), называются *инвариантными*, а по критерию (5.7) — *автономными*.

Если экстремальное значение функционала качества зависит от многих переменных $Q(x_1, x_2, \dots, x_n)$, то устанавливается одна, наиболее существенная из этих переменных x_j , и оптимизация системы осуществляется по критерию оптимальности

$$J = \int_{t_1}^{t_2} \frac{\partial Q(x_1, x_2, \dots, x_n)}{\partial x_j} \Delta x_j(t) dt \rightarrow \min. \quad (5.8)$$

В динамических системах автоматического управления часто ставится задача минимизации расхода энергии, затрачиваемой на функционирование системы. В этом случае минимизируется функционал качества, характеризующий расход энергии. Например, при управлении объектом, питающимся от источника электрической энергии, система, оптимальная по расходу энергии, должна удовлетворять условию

$$J = \int_{t_1}^{t_2} u i dt \rightarrow \min,$$

где u, i — напряжение и ток нагрузки.

В теории управления используются так называемые *минимаксные критерии оптимальности*, характеризующие условия наилучшей работы системы в наихудших возможных условиях. Примером использования минимаксного критерия может быть выбор на его основе варианта системы автоматического управления, имеющей минимальное значение максимального перерегулирования.

Как отмечалось выше, критерий оптимальности реализуется при наличии ограничений, накладываемых на переменные и на показатели качества управления. В системах автоматического управления ограничения, накладываемые на координаты управления, можно подразделить на естественные и условные.

Естественные ограничения отражают ограничения координат по модулю, вытекающие из естественных свойств системы и ее элементов: $|x_i| \leq x_{\max}$. Эти ограничения часто обуславливаются прин-

ципом работы объекта и элементов регулятора. Например, насыщение в электромагнитных цепях ограничивает выходную величину независимо от роста входной, расход топлива ограничен максимальным открытием клапанов и т. д.

Условные ограничения вводятся в систему сознательно с целью обеспечения конкретных требований. Например, сила тока электродвигателя ограничивается по условиям допустимого нагрева при заданном сроке эксплуатации. Условные ограничения обычно удовлетворяются ограничениями, накладываемыми на переменные управления $|u| \geq u_{\max}$, что обеспечивает ограничение всех координат системы:

$$\left| a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_1 \frac{dx}{dt} + a_0 x \right| \leq k |u_m|.$$

Оптимальные системы классифицируются по различным признакам. Отметим некоторые из них.

В зависимости от реализуемого критерия оптимальности различают:

1) системы, оптимальные по быстродействию. Они реализуют критерий минимального времени переходных процессов;

2) системы, оптимальные по точности. Они формируются по критерию минимума отклонения переменных за время переходных процессов или по критерию минимума среднеквадратичной ошибки;

3) системы, оптимальные по расходу топлива, энергии и т. д., реализующие критерий минимума расхода;

4) системы, оптимальные по условиям инвариантности. Они синтезируются по критерию независимости выходных переменных от внешних возмущений или от других переменных;

5) оптимальные экстремальные системы, обуславливающие критерий минимума отклонения показателя качества от его экстремального значения.

В зависимости от характеристик объектов оптимальные системы подразделяются на линейные, нелинейные, непрерывные, дискретные, аддитивные, параметрические. Эти признаки, кроме двух последних, не нуждаются в пояснениях. В аддитивных системах воздействия на объект не изменяют его характеристик. Если же воздействия изменяют коэффициенты уравнений объекта, то такие системы называются *параметрическими*.

В зависимости от типа критерия оптимальности оптимальные системы подразделяются на следующие:

1) равномерно-оптимальные, в которых каждый отдельный процесс протекает оптимально;

2) статистически оптимальные, реализующие критерий оптимальности, имеющий статистический характер из-за случайных воздействий на систему. В этих системах наилучшее поведение обеспечивается не в каждом отдельном процессе, а лишь в некоторых. Статистически оптимальные системы можно назвать оптимальными в среднем;

3) минимаксно-оптимальные, которые синтезируются из условия минимаксного критерия, обеспечивающего лучшим наихудший результат по сравнению с подобным наихудшим результатом в любой другой системе.

По степени полноты информации об объекте оптимальные системы подразделяются на системы с полной и неполной информацией. В состав информации об объекте включаются сведения: о зависимости между входными и выходными величинами объекта; характере возмущения; состоянии объекта; задающем воздействии, определяющем требуемый режим работы системы; цели управления — функционале, выражающем критерий оптимальности.

Информация об объекте в действительности всегда неполная, однако во многих случаях это не оказывает существенного влияния на функционирование системы по избранному критерию оптимальности. В ряде же случаев неполнота информации настолько существенна, что при решении задач оптимального управления требуется использование статистических методов.

В зависимости от полноты информации об объекте управления критерий оптимальности может быть выбран «жестким» (при достаточно полной информации) или «приспосабливающимся», т. е. изменяющимся при изменении информации. По этому признаку оптимальные системы подразделяются на системы с жесткой настройкой и адаптивные. В число адаптивных систем входят экстремальные, самонастраивающиеся и обучающиеся системы. Эти системы наиболее полно отвечают современным требованиям, предъявляемым к системам оптимального управления.

5.2. ЗАДАЧИ СИНТЕЗА ОПТИМАЛЬНЫХ СИСТЕМ И МАТЕМАТИЧЕСКИЕ МОДЕЛИ ОБЪЕКТОВ УПРАВЛЕНИЯ

Решение задачи синтеза оптимальной системы заключается в разработке системы управления, отвечающей заданным требованиям, т. е. в создании системы, реализующей выбранный критерий оптимальности. В зависимости от объема сведений о структуре системы управления задача синтеза ставится в одной из двух следующих постановок.

Первая постановка охватывает случаи, когда структура системы известна. В таких случаях объект и регулятор могут быть описаны соответствующими передаточными функциями, а задача синтеза сводится к определению оптимальных значений числовых параметров всех элементов системы, т. е. таких параметров, которые обеспечивают реализацию выбранного критерия оптимальности.

Во второй постановке задача синтеза ставится при неизвестной структуре системы. В этом случае требуется определить такую структуру и такие параметры системы, которые обеспечат систему, оптимальную по принятому критерию качества. В инженерной практике задача синтеза в такой постановке встречается редко. Чаще всего объект управления либо задан как физическое устрой-

ство, либо описан математически, и задача синтеза сводится к синтезу оптимального регулятора. Следует подчеркнуть, что и в этом случае необходим системный подход к синтезу системы оптимального управления. Суть такого подхода заключается в том, что при синтезе регулятора рассматривается вся система (регулятор и объект) как единое целое.

На начальной стадии синтеза оптимального регулятора задача сводится к его аналитическому конструированию, т. е. к определению его математического описания. При этом одну и ту же математическую модель регулятора можно реализовать различными физическими устройствами. Выбор конкретной физической реализации аналитически определенного регулятора осуществляется с учетом условий работы конкретной системы автоматического управления. Таким образом, задача синтеза оптимального регулятора неоднозначна и может быть решена различными путями.

При синтезе системы оптимального управления весьма важно создание модели объекта, максимально адекватной реальному объекту. В теории управления так же, как в других современных областях науки, основными видами моделей объектов являются математические модели — уравнения статики и динамики объектов.

При решении задач синтеза оптимальной системы единой математической моделью объектов управления обычно является модель в форме уравнений состояния. Под *состоянием системы автоматического управления* в каждый момент времени понимается минимальный набор переменных (переменных состояния), который содержит количество информации, достаточное для определения координат системы в текущем и будущем состояниях системы.

Исходные уравнения объекта обычно нелинейны. Для приведения их к форме уравнений состояний широко используются методы линейных преобразований исходных уравнений.

Для характеристики объекта управления введем следующие понятия:

$X\{x_1, x_2, \dots, x_n\}$ — вектор состояния объекта;

$Y\{y_1, y_2, \dots, y_m\}$ — вектор выходных величин;

$U\{u_1, u_2, \dots, u_r\}$ — вектор управления, прикладываемого к объекту;

$F\{f_1, f_2, \dots, f_l\}$ — вектор возмущения;

$X(t_0) = X_0$ — вектор начального состояния объекта.

Задача оптимизации в таком случае сводится к удовлетворению функционала качества $J[X, U, F]$ при ограничениях $X \in \Omega_X$, $U \in \Omega_U$, где Ω_X — область допустимых значений координат состояния объекта; Ω_U — область допустимых значений управлений.

Широкий класс объектов оптимальных систем в результате линеаризации исходных уравнений может быть описан уравнениями:

$$\dot{X} = f[X, U, F(t)]; \quad Y = \varphi(X, U); \quad X(t_0) = X_0;$$

$$X \in \Omega_X; \quad U \in \Omega_U.$$

Например,

$$\dot{X} = AX + BU + CF(t); \quad Y = DX + EU; \quad X(t_0) = X_0;$$

$$X \in \Omega_X; \quad U \in \Omega_U,$$

где A, B, C, D, E — постоянные коэффициенты.

В теории оптимального управления широко используются следующие математические методы: принцип максимума; динамического программирования; вариационного исчисления; математического программирования. Каждый из перечисленных методов имеет свои особенности и, следовательно, свою область применения.

Принцип максимума позволяет сравнительно легко учесть ограничения на управляющие воздействия, подводимые к объекту управления. Метод наиболее эффективен при синтезе систем, оптимальных по быстродействию. Однако реализация метода даже с использованием ЭВМ значительно затруднена.

Метод динамического программирования обладает большими возможностями. Однако для систем высокого порядка (выше четвертого) использование метода весьма затруднительно. При нескольких переменных управления реализация метода динамического программирования на ЭВМ требует объемов памяти, превышающих возможности современных машин.

Вариационное исчисление применяется при отсутствии ограничений на переменные состояния и на переменные управления. Получение численного решения на базе методов вариационного исчисления затруднительно. Метод используется, как правило, для некоторых весьма простых случаев.

Методы математического программирования (линейного, нелинейного и др.) широко применяются для решения задач оптимального управления как в автоматических, так и в автоматизированных системах. Общая идея методов заключается в отыскании экстремума функции в пространстве многих переменных при ограничениях в виде системы равенств и неравенств. Методы позволяют найти численное решение широкого круга задач оптимального управления.

Достоинствами методов математического программирования является возможность сравнительно просто учитывать ограничения на управления и переменные состояния, а также обычно допустимые требования к объему памяти.

В настоящей главе рассматриваются принцип максимума и метод динамического программирования. Методы математического программирования рассматриваются в гл. 7, посвященной проблемам управления сложными — организационно-техническими и экономическими — системами.

5.3. МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Принцип максимума Понтрягина

Принцип максимума был сформулирован советским ученым Л. С. Понтрягиным в 1956 г. и доказан в ряде его последующих работ, созданных совместно с В. Г. Болтянским и Р. В. Гамкре-

лидзе. Этот принцип достаточно общий для определения условий существования оптимального управления широким кругом систем.

Доказательство принципа максимума базируется на теории множеств, функциональном анализе и представляет определенные затруднения. Поэтому, не останавливаясь на доказательстве рассмотрим методику использования принципа максимума для синтеза оптимальных управлений.

Для использования принципа максимума уравнения объекта приводим к виду

$$\frac{dx_i}{dt} = F_i(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r), \quad i = \overline{1, n}, \quad (5.9)$$

где r — координаты управления,

или в векторной форме $\frac{dX}{dt} = F(X, U)$.

Формулируем функционал качества управления

$$J = \int_{t_1}^{t_2} F(X, U) dt, \quad (5.10)$$

вычисляем функцию Гамильтона

$$H = \sum_{i=1}^n \Psi_i(t) F_i(X, U), \quad (5.11)$$

где $\Psi(t)$ — вспомогательная функция, определяемая на базе уравнений Гамильтона

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial \Psi_i}; \quad - \frac{\partial H}{\partial x_i} = \frac{d\Psi_i}{dt}. \quad (5.12)$$

Принцип максимума утверждает следующие условия оптимального управления: для осуществления оптимального управления необходимо существование таких ненулевых непрерывных функций $\Psi_i(t)$, чтобы функция Гамильтона H как функция переменных управления достигла максимума. Вопрос о существовании функций $\Psi_i(t)$ решается в каждом конкретном случае с учетом физического смысла процессов, протекающих в системе, и состава ее элементов.

Принцип максимума является методом решения нелинейных задач оптимального управления, учитывающий ограничения на управляющие воздействия.

Рассмотрим использование принципа максимума для синтеза оптимальных управлений объектом, имеющим передаточную функцию $W_0(p) = x/u = 1/p^2$, где x — выходная величина объекта управления; u — управление, прикладываемое к объекту.

Из выражения передаточной функции объекта имеем $\frac{d^2x}{dt^2} = u$, ограничения $|u| \leq u_{\max}$.

Введя обозначения $x = x_1$; $\frac{dx}{dt} = x_2$, уравнения объекта приведем к виду (5.9)

$$\frac{dx_1}{dt} = x_2; \quad \frac{dx_2}{dt} = u. \quad (5.13)$$

На основании (5.11) и (5.13) запишем функцию Гамильтона

$$H = \Psi_1 x_2 + \Psi_2 u. \quad (5.14)$$

Используя (5.12) и (5.14), найдем

$$\frac{dx_1}{dt} = \frac{\partial H}{\partial \Psi_1} = x_2; \quad \frac{dx_2}{dt} = \frac{\partial H}{\partial \Psi_2} = u,$$

убедившись в соответствии уравнений Гамильтона функциям основных переменных.

На основании (5.12) и (5.14) получим

$$\frac{d\Psi_1}{dt} = -\frac{\partial H}{\partial x_1} = 0; \quad \frac{d\Psi_2}{dt} = -\frac{\partial H}{\partial x_2} = -\Psi_1. \quad (5.15)$$

Интегрируя (5.15), найдем вспомогательные переменные $\Psi_1 = c_1$; $\Psi_2 = -\int c_1 dt = c_2 - c_1 t$, следовательно, на основании (5.14) имеем:

$$H = c_1 x_2 + (c_2 - c_1 t) u = c_1 \frac{dx}{dt} + (c_2 - c_1 t) u. \quad (5.16)$$

В соответствии с принципом максимума оптимальное управление будет обеспечено, если (5.16) принимает максимальное положительное значение, т. е. второе слагаемое должно быть положительным и наибольшим. Это условие соблюдается, если управление обеспечивается по закону

$$u = U_{\max} \operatorname{sign} \Psi_2, \text{ т. е. } u = U_{\max} \operatorname{sign} (c_2 - c_1 t). \quad (5.17)$$

Оптимальная по быстродействию система управления обеспечивает на фазовой плоскости x_2 ; x_1 наиболее быстрый переход изображающей точки из произвольного начального состояния в начало координат, где $x_i = 0$; $\frac{dx_i}{dt} = 0$. Так как в (5.17) функция $c_2 - c_1 t$, являясь линейной функцией времени, может только один раз изменить свой знак, то указанный характер движения изображающей точки на фазовой плоскости возможен лишь при двух значениях функции управления u : $+u_{\max}$; $-u_{\max}$.

В уравнениях (5.13) примем $u=1$ и исключим время t , после чего получим $dx_1 = x_2 dx_2$. Интегрируя, находим:

$$x_1 = x_2^2/2 + c \text{ при } u = +1; \quad (5.18)$$

$$x_1 = -x_2^2/2 + c \text{ при } u = -1. \quad (5.19)$$

Выражения (5.18) и (5.19) являются уравнениями фазовых траекторий рассматриваемой системы, оптимальной по быстродействию. При $c=0$ выражение (5.18) описывает параболу BOB_1 , проходящую через начало координат, а (5.19) — параболу AOA_1 , также проходящую через начало координат (рис. 5.1). При $c \neq 0$ выражения (5.18) и (5.19) описывают множество парабол, сдвинутых вправо и влево от оси ординат. В синтезируемой оптимальной сис-

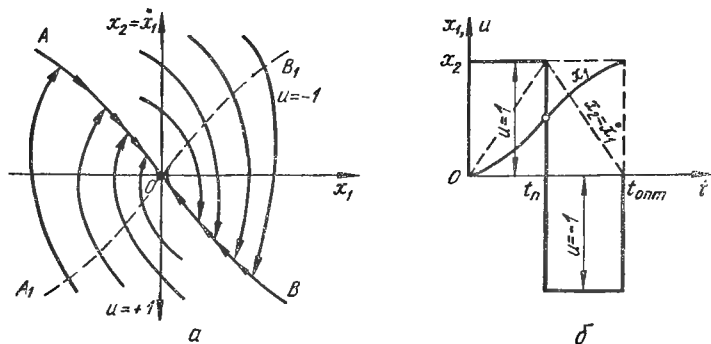


Рис. 5.1. Фазовый портрет системы, оптимальной по быстродействию:
 а — фазовые траектории; б — графики изменения фазовых координат и управляющего воздействия во времени

теме должны быть реализованы только те фазовые траектории, которые при управляющих воздействиях $u = \pm 1$ обеспечивают быстрое движение изображающей точки к началу координат. При $c=0$ и $u=1$ этому условию отвечает траектория BO , а при $u=-1$ — траектория AO . Следовательно, $A0B$ является линией, определяющей моменты переключения знака управления для обеспечения оптимального переходного процесса. Аналитическое выражение линии переключения можно получить из уравнений (5.18) и (5.19), приняв $c=0$:

$$x_{2п} = \left(\frac{dx_1}{dt} \right)_п = (\text{sign } x_1) \sqrt{2|x_1|}. \quad (5.20)$$

Очевидно, что для осуществления управления, оптимального по быстродействию, необходимо обеспечивать изменение управляющего воздействия, прикладываемаемого к объекту управления, по закону

$$u = \begin{cases} +1 & \text{при } \frac{dx}{dt} < \left(\frac{dx}{dt} \right)_п \text{ и при } \frac{dx}{dt} = \left(\frac{dx}{dt} \right)_п \\ & \text{для всех } \left(\frac{dx}{dt} \right)_п < 0; \\ -1 & \text{при } \frac{dx}{dt} > \left(\frac{dx}{dt} \right)_п \text{ и при } \frac{dx}{dt} = \left(\frac{dx}{dt} \right)_п \\ & \text{для всех } \left(\frac{dx}{dt} \right)_п > 0. \end{cases} \quad (5.21)$$

На рис. 5.1 вместе с фазовым портретом представлены графики изменения фазовых координат системы и изменения управляющего воздействия, обеспечивающего управление, оптимальное по быстродействию. Время переключения $t_п$ можно определить из (5.17) в момент, когда $u=0$: $t_п = c_2/c_1$.

Уравнения (5.21) выражают результаты аналитического конструирования регулятора системы, оптимальной по быстродействию, имеющей объект управления с передаточной функцией $W_0(p) = 1/p^2$. Очевидно, что физическая реализация системы связана с необходимостью использования переключающего реле, обеспечивающего переключение с $u = +1$ на $u = -1$ в момент времени $t = t_n$, и вычислительного устройства для вычисления $\frac{dx}{dt}$ и сравнения этой величины с величиной $\left(\frac{dx}{dt}\right)_n$.

Метод динамического программирования (метод Беллмана)

Метод широко применяется для исследования оптимального управления как в динамических (технических), так и в экономических системах. Для реализации метода динамического программирования связи в системе между выходными переменными, управлениями и критериями оптимальности могут быть заданы как в виде аналитических зависимостей, так и в виде таблиц численных данных, экспериментальных графиков и т. п.

В основе метода динамического программирования лежит сформулированный Р. Беллманом «принцип оптимальности». Для формулировки принципа Беллмана назовем управление, обеспечивающее оптимальное движение в системе, т. е. оптимальное управление $u(t)$, *решением*. Принцип оптимальности Беллмана утверждает: оптимальное поведение системы характеризуется тем свойством, что каковы бы ни были первоначальное состояние и решения до некоторого момента времени, последующие решения должны составлять оптимальное поведение относительно состояния, получающегося в результате первого решения.

Принцип Беллмана вытекает из следующих рассуждений (рис. 5.2). Допустим, что в результате оптимального управления изображающая точка в фазовой плоскости переменных системы x_2, x_1 движется по траектории abc из начального состояния $x(0)$ в конечное состояние $x(T)$. Допустим, что процесс движения начали рассматривать не с момента $t=0$, т. е. не с точки $a, x(0)$, а с некоторого момента $t=t_1$, т. е. с точки $b, x(t_1)$. Очевидно, что дальнейшее управление будет оптимальным, если движение изображающей точки обеспечится по траектории bec , т. е. по той части траектории, которая была определена первоначальным решением, обеспечившим движение по траектории $abec$. Действительно, пусть, начиная с момента времени $t=t_1$, определена другая траектория, отличающаяся от траектории bec , например bdc , показанная на рис. 5.2. штриховой линией. Эта траектория обес-

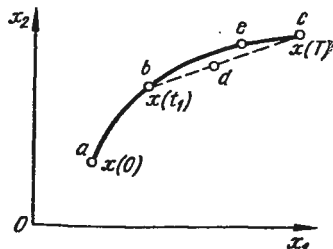


Рис. 5.2. Принцип оптимальности Беллмана

печивает меньшее значение функционала качества, выражающего потери на участке $x(t_1) \rightarrow x(T)$. Однако в таком случае окажется, что предыдущее решение, определившее оптимальную траекторию *abes*, было не оптимальным, что противоречит условию.

Для использования метода динамического программирования уравнения объекта нужно привести к виду

$$\frac{dx}{dt} = f(x, u). \quad (5.22)$$

Должен быть задан функционал, достигающий минимума при оптимальном управлении,

$$J = \int_0^T F(x, u) dt \quad (5.23)$$

и ограничения на управляемую и управляющую переменные: $x \in X$; $u \in \Omega$.

Весь интервал времени управления $0 - T$ разобьем на N одинаковых промежутков

$$\Delta t = T/N. \quad (5.24)$$

Дифференциальное уравнение (5.22) заменим разностным уравнением

$$\frac{dx}{dt} \approx \frac{\Delta x}{\Delta t} = \frac{x_{i+1} - x_i}{\Delta t} = f(x_i, u_i), \quad (5.25)$$

где

$$x_1 = x(i\Delta t), \quad i = \overline{1, N}; \quad x_{i+1} = x_i + \Delta t f(x_i, u_i).$$

Переход от непрерывных функций (5.22) к разностным уравнениям влечет за собой замену функционала качества (5.23) приближенной суммой:

$$J \approx \Delta t \sum_{i=0}^{N-1} F(x_i; u_i). \quad (5.26)$$

Верхний предел суммы $N - 1$ объясняется тем, что площадь, выражаемая интегралом (5.23), заменяется суммой элементарных площадей с левой ординатой.

Рассмотрим процесс управления на последнем интервале времени, т. е. в интервале $(N-1)\Delta t \leq t \leq N\Delta t = T$.

Допустим, что значение управляемой переменной x_{N-1} в начале последнего интервала известно. Очевидно, что оптимальное управление на последнем интервале должно минимизировать функционал вида

$$J_{N-1} = \Delta t F(x_{N-1}, u_{N-1}), \quad (5.27)$$

представляющий собой последний член в сумме (5.26). Следовательно, решение, т. е. закон оптимального управления на последнем этапе u_{N-1} , обеспечивая $J_{N-1} \rightarrow \min$, должен удовлетворить заданным ограничениям.

Таким образом, найдено оптимальное управление на последнем участке Δt как функция управляемой переменной x_{N-1} в начале интервала $u_{N-1} = \Psi_1(x_{N-1})$.

Рассуждая аналогично, для предпоследнего интервала $(N-2)\Delta t \leq t \leq (N-1)\Delta t$ найдем оптимальное управление

$$u_{N-2} = \Psi_2(x_{N-2}).$$

Приведенные рассуждения справедливы для всех интервалов $N\Delta t \leq t \leq 0$, т. е. $(N-i)\Delta t \leq t \leq (N-i+1)\Delta t$; $u_{N-i} = \Psi_i(x_{N-i})$. Значения u_{N-i} определяем после того, как составлены уравнения для всех интервалов Δt , включая первый, на котором $x(0)$ задано. Таким образом, сложная задача поиска оптимального управления на всем времени управления заменяется более простой задачей нахождения оптимального управления на каждом шаге Δt .

В случае многомерной системы уравнение (5.22) и функционал (5.23) становятся векторными, т. е. имеем:

$$\frac{dx}{dt} = F(X, U); \quad X\{x_1, x_2, \dots, x_n\}; \quad U\{u_1, u_2, \dots, u_r\};$$

$$J = \int_0^T \Psi(X, U) dt,$$

где n — число управляемых координат; r — число управлений.

Алгоритм вычисления оптимальных управлений в многомерной системе аналогичен рассмотренному выше, но на каждом шаге Δt должна осуществляться минимизация частичного функционала J_{N-1} по всем r переменным управления u_r .

Как видно из изложенного, численные значения управляемых переменных $x(t)$ и управлений $u(t)$ по методу Беллмана получаются в дискретные моменты времени. Следовательно, точность решения зависит от того, насколько разностные уравнения эквивалентны соответствующим непрерывным уравнениям. Для систем выше второго порядка замена дифференциальных уравнений разностными приводит, как правило, к приближенным решениям задачи.

В задачах оптимального управления принцип оптимальности Беллмана выражается в виде нелинейного уравнения в частных производных первого порядка, полученного в результате применения метода динамического программирования к функции, выражающей оптимальное значение функционала в зависимости от начального состояния.

Пусть интервал управления $0 - T$ не фиксирован, тогда $J(x, u)$ будет функцией T . Если $J(x, u)$ является непрерывной функцией всех переменных и всюду имеет непрерывные частные производные, то в интервале $t, t + \Delta t$ имеем:

$$J[x(t), T] = \min_{u \in \Omega_u} \left\{ \int_t^{t+\Delta t} F(x, u) dt + J[x(t + \Delta t), T] \right\}.$$

Разложив интеграл в ряд Тейлора по степеням Δt в окрестности t и перейдя к пределу при $\Delta t \rightarrow 0$, получим уравнение Беллмана

$$\frac{\partial J}{\partial t} = \min_{u \in \Omega_u} \left[F(x, u, t) + \sum_{i=1}^n \frac{\partial J(x, t)}{\partial x_i} f_i(x, u, t) \right] \quad (5.28)$$

Непосредственное нахождение оптимального управления на базе уравнения Беллмана весьма затруднительно и возможно лишь в ограниченных частных случаях. Тем не менее уравнение Беллмана позволяет обосновать некоторые варианты метода динамического программирования, обеспечивающие получение приближенных решений.

Обратимся к уравнению Беллмана (5.28). Если J в явном виде не зависит от t , то $\frac{\partial J}{\partial t} = 0$. В задаче об управлении, оптимальном по быстродействию, $F(x, u) \equiv 1$. В результате уравнение Беллмана для систем, оптимальных по быстродействию, принимает вид

$$\min_{u \in \Omega_u} \left[\sum_{i=1}^n \frac{dJ}{dx_i} f_i(x, u) \right] = -1.$$

5.4. ЭКСТРЕМАЛЬНЫЕ АВТОМАТИЧЕСКИЕ СИСТЕМЫ УПРАВЛЕНИЯ

Развитие техники управления потребовало создания таких автоматических систем, которые должны обладать свойством приспособляемости к изменениям внутренних и внешних условий функционирования. В таких системах параметры могут изменяться в широких пределах, причем сведения об этих изменениях обычно недостаточны для формирования управляющих воздействий на основании текущего измерения переменных состояния системы. Рассмотренные выше принципы управления (по отклонению, по возмущению, комбинированные) не только не обеспечивают требуемого качества управления, но в ряде случаев оказываются несостоятельными для создания работоспособной системы. В связи с этим во второй половине текущего столетия началась разработка и широкое внедрение адаптивных систем, способных приспособляться к изменениям внешних и внутренних условий (*adaptio*, лат.— приспособление). Адаптация в таких системах может рассматриваться как процесс накопления и использования информации в системе, направленный на достижение ее оптимального состояния при начальной неопределенности и изменяющихся условиях.

Адаптация применяется тогда, когда факторы, воздействующие на систему, полностью или частично неизвестны. В процессе адаптации система накапливает сведения об этих факторах и определяет их характеристики.

В настоящее время разработано множество типов адаптивных систем, среди которых широкое распространение в различных областях техники нашли экстремальные системы (ЭС) автоматического управления. Эти системы содержат объект с экстремальной статической характеристикой и оптимизатор — автоматическое устройство, обеспечивающее работу объекта в экстремальном ре-

жиме. При этом под экстремальным режимом понимается работа системы в районе экстремума статической характеристики.

В ЭС формируются такие управления на входе объекта (на выходе регулятора), которые обеспечивают экстремум функционала качества $J=G(X, F, U)$, где X — вектор управляемых переменных; F — вектор возмущающих воздействий; U — вектор управлений.

ЭС наиболее эффективны в случаях, когда объект обладает статической характеристикой с достаточно ярко выраженным экстремумом, т. е. существует показатель качества управления, с изменением которого достигается существенное изменение технико-экономической эффективности. Техническая реализация принципа оптимального управления связана с необходимостью определения экстремума функционала качества и соответствующего воздействия на регулирующие органы объекта. Поэтому решение об использовании экстремального управления принимается после анализа его технико-экономической эффективности с учетом дополнительных, иногда значительных затрат на обеспечение экстремального управления.

В настоящее время налажено промышленное изготовление оптимизаторов — типовых экстремальных регуляторов для использования в различных областях техники управления с целью существенного повышения технико-экономических показателей производственных и технологических процессов.

Примером ЭС может служить система регулирования температуры в газовых печах обжига строительных материалов (рис. 5.3, а). Интенсивность горения топлива и, следовательно, интенсивность нагрева X в печи определяется соотношением количеств подаваемого топлива q и воздуха v . На рис. 5.3, б показаны экстремальные статические характеристики объекта $X=f(v)$ при различных расходах подаваемого топлива. Оптимальный режим работы определяется точкой A , лежащей на экстремуме характеристики $X=f(v)$. Форма кривой $X(v)$ и положение максимума зависят от многих факторов (возмущений) — расхода топлива, его калорийности, температуры подаваемого воздуха и ряда других. Задача ЭС в данном случае заключается в том, чтобы в любой обстановке, в пределах заданных технических условий обеспечить оптимальный режим работы системы, т. е. режим, характеризуемый точкой A статической характеристики объекта $X(v)$. Это обеспечивается соответствующим регулированием подачи воздуха при различных

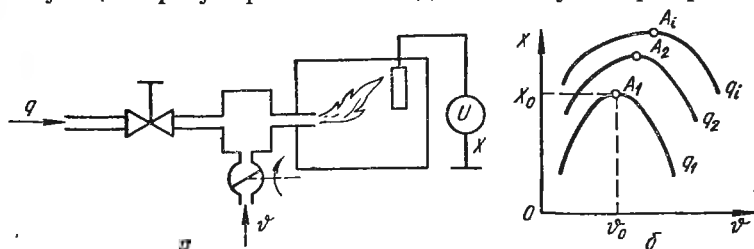


Рис. 5.3. Система экстремального регулирования:

а — схема системы; б — статические характеристики объекта управления

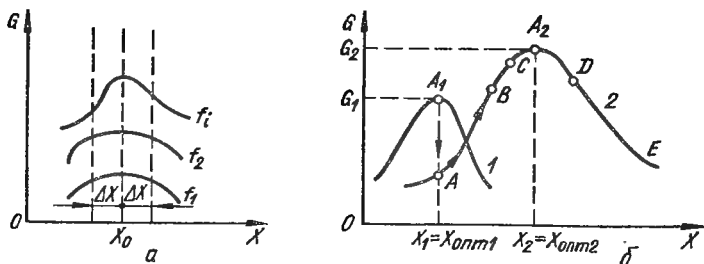


Рис. 5.4. Статические характеристики объектов экстремальных систем:
а — с постоянным экстремумом; б — с изменяющимся экстремумом

условиях работы системы, определяемых указанными возмущениями.

В общем случае экстремальная автоматическая система управления должна выводить объект на экстремальный режим и удерживать его в нем. Сложность ЭС в значительной степени зависит от свойств объекта, выраженных в форме и положении его статических характеристик.

В наиболее простых случаях экстремальные характеристики объектов стабильны (фиксированы), и оптимальный режим достигается формированием управлений по отклонению управляемой величины от ее экстремального значения, а сами системы называются *статическими экстремальными*. ЭС подобного типа аналогичны рассмотренным ранее системам автоматической стабилизации регулируемой величины.

Более характерным типом ЭС являются динамические экстремальные системы, в которых статические характеристики объектов не фиксированы. Их положение может определяться как детерминированными факторами, так и случайными, причем второй тип систем является наиболее общим.

Для простоты рассуждений сначала остановимся на динамических ЭС с одномерным объектом и экстремальным показателем качества $G(X, f)$, где X — регулируемая величина; f — возмущение. Рассмотрим два типа статических характеристик объекта — с экстремумом при фиксированном значении X (рис. 5.4, а) и с экстремумом, изменяющимся произвольно (рис. 5.4, б).

В первом случае уход объекта от оптимального режима сопровождается возникновением отклонения $\Delta X = X_0 - X$, где X_0 — значение регулируемой величины в экстремальном режиме; X — текущее значение регулируемой величины. В этом случае ЭС строится как обычная система стабилизации регулируемой переменной с использованием принципа регулирования по отклонению.

Во втором случае задача обеспечения экстремального управления существенно усложняется, так как в зависимости от различных факторов экстремум характеристики $G(X, f)$ смещается как по вертикали, так и по горизонтали. В таких случаях для осуществления экстремального управления необходимо автоматическое слежение за положением точки экстремума характеристики объекта.

Пусть в некоторый момент времени состояние системы характеризуется оптимальным режимом, определяемым точкой на статической характеристике 1 (рис. 5.4, б). В этом режиме $X = X_1 = X_{\text{опт1}}$; $G = G_1 = G_{\text{опт1}}$. Допустим, что под влиянием возмущений статическая характеристика $G(X, f)$ объекта переместилась из положения 1 в положение 2. Тогда при отсутствии регулирования система перейдет из оптимального режима, характеризуемого точкой A_1 , в режим, характеризуемый точкой A , лежащей на экстремальной характеристике 2. Очевидно, что такой режим не является оптимальным. Для приведения системы в оптимальное состояние в ней должны быть предусмотрены устройства, способные определить новое положение экстремума (точки A_2) и изменить координату X со значения X_1 до значения X_2 .

Рассмотренная картина определяет принципы построения систем экстремального управления и техническую базу для реализации этих принципов.

Задача экстремального управления упрощается, если известен закон перемещения статической характеристики во времени. В таких случаях экстремальное состояние системы можно обеспечить с помощью программного управления.

При неизвестном законе изменения статической характеристики объекта ЭС строится как автоматическая поисковая система, обеспечивающая автоматическое слежение за положением экстремума статической характеристики. В такой системе осуществляются две операции, приводящие систему в экстремальное состояние: 1) поисковая (пробная) операция; 2) рабочая операция. В результате поисковой операции выявляется соотношение между текущим значением показателя качества управления $J = G(X, f, u)$ и его экстремальным значением. В результате рабочей операции изменяется управляющее воздействие по закону, обеспечивающему перевод объекта в оптимальный (экстремальный) режим работы.

Для определения требуемого закона изменения управления сравниваются два соседних значения показателя качества $G(X)$, отличающихся приращением аргумента X . Пусть исходное состояние системы характеризуется значением $G(X_1)$. Изменяя сигнал управления, найдем $G(X_2) = G(X_1 + \Delta X)$. Если $G(X_2) > G(X_1)$, то выбранное направление изменения управляющего воздействия способствует движению системы к экстремуму-максимуму. Если $G(X_2) < G(X_1)$, то для приведения объекта в экстремальный режим необходимо изменить приращение сигнала управления на противоположный. Приведенные рассуждения относятся к объектам, в статических характеристиках которых имеется экстремум-максимум. При статических характеристиках, содержащих экстремум-минимум, характер изменения управляющего воздействия будет противоположным.

Следовательно, поиск экстремума связан с определением крутизны статической характеристики объекта в точке, соответствующей данному режиму работы, т. е. с определением производной

$$S = \left[\frac{dG(X)}{dX} \right]_{X=X_0}$$

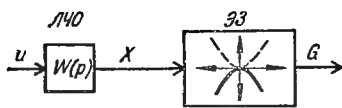


Рис. 5.5. Функциональная схема объекта управления в экстремальной системе

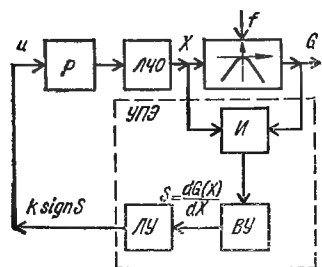


Рис. 5.6. Экстремальная система с определением знака производной функционала качества

где X_0 — значение регулируемой переменной в текущий момент. Очевидно, что система окажется в экстремальном режиме при условии $S=0$.

Описанный способ поиска экстремума по производной показателя качества получил название *поиска экстремума по чувствительности*. В практике ЭС широко применяются и другие методы поиска экстремума, например по знаку приращения ΔG , по экстремуму и др.

Структура системы экстремального управления в значительной степени зависит от принятого и реализуемого способа поиска экстремума. По этому признаку ЭС подразделяются на следующие типы: 1) с определением производной показателя качества; 2) с определением знака производной; 3) шаговые экстремальные; 4) с модулирующим поисковым сигналом; 5) с запоминанием экстремума.

После выхода на экстремальный режим ЭС функционирует как система регулирования по отклонению, в которой заданным значением регулируемой величины является ее значение, обеспечивающее экстремум функционала качества.

Структура одномерного объекта может быть представлена в виде последовательного соединения двух звеньев (рис. 5.5) — линейной части объекта (ЛЧО) с передаточной функцией $W(p)$ и экстремального звена (ЭЗ) со статической характеристикой, содержащей экстремум-максимум или экстремум-минимум.

На рис. 5.6 в качестве примера приведена функциональная схема системы экстремального регулирования, в которой экстремум статической характеристики объекта находится путем определения знака производной функционала качества $G(X, f)$. В этой системе поиск экстремума основан на том, что знак производной функции $G(X, f)$ однозначно определяет положение точки на статической характеристике объекта. Действительно, если статическая характеристика объекта имеет экстремум-максимум, то положительное значение производной означает, что точка находится на левой ветви характеристики, а отрицательное значение производной определяет точку, находящуюся на правой ветви характеристики (в случае экстремума-минимума — наоборот). В связи с этим ЭС снабжена устройством поиска экстремума (УПЭ), включающим в себя идентификатор (И) и вычислительное устройство (ВУ). Идентификатор предназначен для определения статической характеристики объекта $G(X, f)$, а вычислительное устройство определяет производную $S = \left[\frac{dG(X, f)}{dX} \right]_{X=X_0}$ и формирует управляющее воздействие, пропорциональное значению производной. Логическое

устройство (ЛУ) определяет знак производной. Управляющее воздействие, формируемое по закону $u = k \text{sign} S$, прикладывается ко входу регулятора (Р).

Остановимся кратко на особенностях построения многомерных систем экстремального управления. (Под многомерными ЭС понимаются системы, реализующие векторный критерий оптимизации, имеющие несколько выходных и несколько входных переменных.)

Общим условием экстремума функции многих переменных является равенство нулю всех ее частных производных по параметрам. На практике экстремальное управление в многомерных ЭС ограничивается тем, что его необходимо обеспечить для одной выходной переменной, а для остальных — заданные граничные условия. В связи с такой постановкой задачи для оптимизации многомерных ЭС используются методы математического программирования, рассмотренные подробно в последующих частях книги.

В многомерных ЭС особое внимание уделяется разработке средств слежения за дрейфующим экстремумом.

Так же, как в одномерных ЭС, в многомерных системах экстремального управления имеют место операции поиска экстремума и рабочие операции, которые могут быть объединены. В последнем случае оптимальные управления определяются в процессе движения к экстремуму, когда каждый рабочий шаг играет роль поискового шага для последующего определения управляющих воздействий.

Многомерные экстремальные системы снабжаются устройствами вычисления градиента и устройствами формирования сигналов управления.

Как известно, градиент — это вектор, показывающий направление наискорейшего изменения некоторой величины, значение которой изменяется от одной точки пространства к другой. Например, для функции $H(x, y, z)$ составляющими градиента будут $\frac{\partial H}{\partial x}$; $\frac{\partial H}{\partial y}$; $\frac{\partial H}{\partial z}$.

В многомерных ЭС в процессе поиска вычисляется градиент $\nabla_u G(X, u)$. Устройства, выполняющие эти вычисления, основываются на различных принципах, зависящих от методов определения частных производных $\frac{\partial G}{\partial u_i}$ и применяемого алгоритма вычислений.

К широко применяемым методам определения частных производных относятся методы конечных приращений, производной по времени, синхронного детектирования и др.

Метод конечных приращений сводится к замене частных производных $\frac{\partial G}{\partial u_i}$ отношениями конечных приращений $\frac{\Delta G}{\Delta u_i}$. В процессе применения метода поочередно изменяются координаты управления u_i и вычисляются соответствующие приращения составляющих градиента $\frac{\Delta G}{\Delta u_i}$.

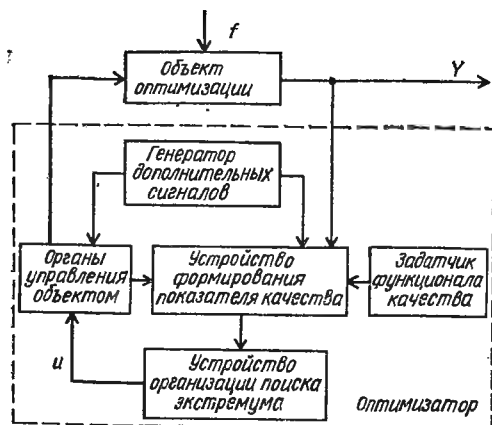


Рис. 5.7. Принципиальная схема многомерной экстремальной системы

поисковыми колебательными сигналами различных частот и амплитуд. При этом количество детекторов равно числу независимых переменных, определяющих экстремум функционала качества $G(X, u)$. Если разложить функцию $G(X, u)$ в ряд Тейлора по степеням приращения Δu_i , вызванных модулирующими колебаниями, то можно убедиться в том, что выходные переменные синхронных детекторов приблизительно пропорциональны частным

производным $\frac{\partial G}{\partial u_i}$. Использование метода синхронного детектирования значительно сокращает время вычисления градиента $\nabla_u G(X, u)$, так как составляющие градиента определяются по соответствующим независимым переменным параллельно. Это становится возможным потому, что модулирующие сигналы разделены по частотному спектру. Частоты поисковых колебаний выбираются (из условия обеспечения амплитуд колебаний на выходе объекта) пропорциональными частным производным $\frac{\partial G}{\partial u_i}$.

Устройство формирования управляющих воздействий в многомерных ЭС содержит дополнительно ряд элементов, в том числе логических, необходимых для изменения переменных управления. Принцип действия устройства определяется методами организации движения системы к экстремуму, к числу которых относятся метод наискорейшего спуска, градиентный, покоординатный и др.

Метод наискорейшего спуска основан на организации поочередного движения системы в направлениях l до тех пор, пока не будет обеспечено условие $\frac{\partial G}{\partial t} = 0$.

Метод градиента основан на изменении производных координат фазового пространства, пропорциональных изменению функционала качества в этих направлениях:

$$\frac{dx_i}{dt} = a \frac{\partial G}{\partial x_i}; \quad \Delta x_i = a \left(\frac{\partial G}{\partial x_i} \right) \Delta t.$$

Метод производной аналогичен методу конечных приращений, отличаясь тем, что вычисляются частные производные $\frac{\partial G}{\partial u_i}$, а не конечные приращения.

Оба метода связаны с поочередными изменениями и вычислениями составляющих градиента, что влечет за собой значительное время определения градиента.

Метод синхронного детектирования основан на модулировании переменных управления дополнительными

Покоординатный метод базируется на известном методе Гаусса — Зайделя. Он связан с организацией движения в направлениях $\frac{\partial G}{\partial x_i} = 0$ и имеет то достоинство, что позволяет использовать принципы построения одномерных экстремальных систем.

На рис. 5.7 представлена принципиальная схема многомерной экстремальной системы.

5.5. ПЕРЕХОДНЫЕ ПРОЦЕССЫ В ЭКСТРЕМАЛЬНЫХ АВТОМАТИЧЕСКИХ СИСТЕМАХ УПРАВЛЕНИЯ

Динамические процессы в экстремальных системах существенно отличаются от процессов в линейных непрерывных системах, в связи с чем требуется применение специальных методов исследования переходных режимов ЭС. Прежде всего, необходимо вводить дополнительные показатели качества управления в ЭС, обусловленные процессами поиска экстремума и работой системы в районе экстремума функционала качества. Такими важными показателями являются показатели качества процессов поиска экстремума и помехоустойчивость.

К основным характеристикам процессов поиска экстремума относятся: устойчивость процесса, сходящегося к окрестностям экстремума; степень приближения к экстремуму; быстродействие, определяемое временем выхода системы на экстремум после смещения экстремальной характеристики объекта.

Устойчивость поискового процесса зависит от структуры и параметров системы, от характера внешних воздействий, приводящих к перемещению экстремальной характеристики. Количественными показателями качества процессов поиска являются время поиска экстремума, амплитуда и период колебаний при движении к экстремуму, потери на поиск экстремума.

Для широкого круга экстремальных систем структурная схема объекта может быть приближенно представлена в виде последовательного соединения двух звеньев (рис. 5.8) — безынерционного нелинейного со статической характеристикой $y=f(x)$ и апериодического с постоянной времени T . В таких случаях уравнение объекта ЭС приводится к виду

$$T \frac{dz}{dt} + z = y = f(x), \quad (5.29)$$

откуда

$$\frac{dz}{dt} = \frac{1}{T} [f(x) - z]. \quad (5.30)$$

Если $f(x)$ имеет экстремум-максимум, а сама функция и ее первая производная непрерывны в интервале $-\infty < x < \infty$, то $f'(x_0) = 0$ и $f''(x_0) < 0$, где x_0 — значение аргумента функции $f(x)$, соответствующее экстремальному значению функции.

Характер движения в экстремальной системе наиболее наглядно представляется в координатах «выход —

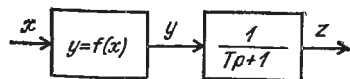


Рис. 5.8. Структурная схема объекта экстремальной системы

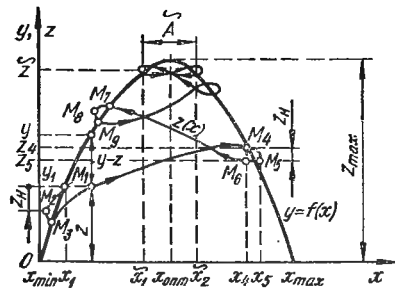


Рис. 5.9. Переходный процесс в экстремальной системе

вход» объекта, а не «выход — время», как было в непрерывных системах автоматического управления. Следовательно, переходный процесс в ЭС следует рассматривать в координатах z, x , т. е. как функцию $z(x)$. Эта зависимость может быть получена из (5.30) исключением времени.

Для определенности примем постоянную скорость перемещения исполнительного механизма, обеспечивающего формирование управляющего воздействия, т. е. $\frac{dx}{dt} = u$;

$u = \pm k$; $k = \text{const}$. Тогда, исключив из (5.30) время, получим

$$\frac{dz}{dx} = \pm \frac{1}{kT} [f(x) - z]. \quad (5.31)$$

Решая (5.31), можно получить зависимость $z(x)$, характеризующую динамику экстремальной системы. Уравнение (5.31) удобно решать методом Эйлера (рис. 5.9).

Пусть момент включения регулятора характеризуется точкой $M_1(x_1, y_1)$ на плоскости $y; x$ или $z; x$. В соответствии с идеей метода Эйлера интегральную кривую $z(x)$, проходящую через точку M_1 , будем искать в виде ломаной линии, состоящей из прямолинейных отрезков, заключенных между ординатами аргументов $x=x_i$ и $x=x_{i+1}$. Угол наклона каждого такого отрезка определится выражением

$$\frac{z_{i+1} - z_i}{\Delta x} = \varphi(z_i, x_i),$$

где $\Delta x = x_{i+1} - x_i$. Кроме того, $\Delta z_i = z_{i+1} - z_i$, следовательно, $\Delta z_i = \Delta x \varphi(z_i, x_i)$.

Наклон отрезка ломаной линии $z(x)$ в точке M_1 (рис. 5.9.) определим согласно (5.31)

$$\Delta z_1 = \frac{\Delta x}{kT} [f(x_1) - z_1],$$

где $f(x_1)$ — значение функции $y=f(x)$ при $x=x_1$.

В точке M_2 выходная величина $z_2 = z_1 + \Delta z_1$.

Наклон ломаной линии $z(x)$ в точке M_2

$$\Delta z_2 = \frac{\Delta x}{kT} [f(x_2) - z_2].$$

Для других точек рассуждения аналогичные.

Направления приращения величины z могут быть определены из рассмотрения условий работы исполнительного механизма, изменяющего переменную x . Действительно, пусть момент включения регулятора в работу $t = t_1$ и характеризуется точкой $M_1(x_1, z_1)$. Предположим при этом, что регулятор включил исполнительный механизм в сторону уменьшения переменной x . Тогда $u = -k$, а в правой части (5.31) будет также минус. Но так как $z > f(x)$, то $\frac{dz}{dx} > 0$ и, следовательно, в точке M_1 уменьшение величины x

сопровождается уменьшением величины z . Это будет продолжаться до тех пор, пока в некоторой точке M_2 разность $z_1 - z_2 = z_H$ достигнет зоны нечувствительности z_H сигнум-реле. После этого исполнительный механизм экстремальной системы реверсируется, переменная x станет увеличиваться. В правой части (5.31) после реверса исполнительного механизма будет знак плюс, а $\frac{dz}{dx} < 0$, так как $f(x) < z$.

В точке M_3 , где $f(x) = z$, согласно (5.31), имеем $\frac{dz}{dx} = 0$, далее до точки $M_4(x_4, z_4)$ производная $\frac{dz}{dx} > 0$ и увеличивается выходная переменная z . В точке M_4 , где $f(x) = z$, имеем вновь $\frac{dz}{dx} = 0$, далее $z > f(x)$, $\frac{dz}{dx} < 0$. После достижения зоны нечувствительности z_H сигнум-реле, т. е. при возникновении условия $z_4 - z_5 = z_H$, исполнительный механизм вновь реверсируется. Далее характер движения повторяется и продолжается до тех пор, пока не установятся автоколебания вокруг экстремального состояния системы. При этом переменная x будет совершать колебательные движения вокруг значения $x = x_{\text{опт}}$.

Автоколебания в ЭС отразятся на плоскости $x; z$ фазовой траекторией в виде петли с координатами $\tilde{z}$, $\tilde{x}_1$, $\tilde{x}_2$, симметричной относительно ординаты $x = x_{\text{опт}}$.

Переходный процесс в ЭС во времени, т. е. функция $x(t)$, может быть получен путем решения уравнения (5.30), выполняемого аналогично. Однако представление динамики ЭС на фазовой плоскости z, x более полно отражает ее свойства.

Как видно из рис. 5.9, в режиме автоколебаний в районе экстремума $\tilde{z} < z_{\text{max}}$, т. е. среднее значение выходной величины объекта меньше экстремального. Это вызвано остановкой исполнительного механизма, обуславливающей «рыскание» изображающей точки вокруг $x = x_{\text{опт}}$. Величина $x_{\text{max}} - \tilde{z} = x_H$ называется *потерей на поиск*. Используя теорему о среднем значении функции, величину x_H можно определить из выражения

$$z_H = z_{\text{max}} - \tilde{z} = z_{\text{max}} - \frac{1}{\tilde{x}_2 - \tilde{x}_1} \int_{\tilde{x}_1}^{\tilde{x}_2} \varphi(x) dx,$$

где $\varphi(x)$ — часть фазовой траектории, образованная одной ветвью петли автоколебаний вокруг экстремума от $x = \tilde{x}_1$ до $x = \tilde{x}_2$.

Потери на поиск — один из показателей качества управления в экстремальных системах. Другой важной характеристикой качества ЭС является амплитуда $\tilde{A}/2$ автоколебаний на входе объекта. Меньшее значение $\tilde{A}/2$ соответствует более качественной работе ЭС. Амплитуда $\tilde{A}/2$ обычно представляется в относительных единицах, в которых за базовые приняты значения $x_{\text{max}} - x_{\text{min}}$.

$$A = \frac{\tilde{A}}{2(x_{\max} - x_{\min})}.$$

Поскольку экстремальные системы являются автоколебательными, то в задачу динамического расчета входит определение параметров автоколебаний: периода автоколебаний выходной величины объекта; амплитуды колебаний входной и выходной величин, среднего уровня $\tilde{z}$ автоколебаний. Исходными данными для расчета параметров автоколебаний являются статическая характеристика $y=f(x)$ объекта управления, постоянная времени объекта T , скорость перемещения выходного вала исполнительного механизма $\frac{dx}{dt} = k$, зона нечувствительности сигнум-реле z_n .

Задача определения параметров автоколебаний экстремальных систем может быть решена либо аналитическим, либо графоаналитическим методом.

Точным аналитическим методом задача решается в таком порядке: 1) составляется и решается дифференциальное уравнение движения в ЭС при произвольных начальных условиях; 2) определяются максимум решения и значения выходных величин, обуславливающих реверс исполнительного механизма; 3) на решение накладывается условие периодичности; 4) определяются параметры периодических движений, которые являются параметрами автоколебаний.

Точные методы исследования автоколебаний ЭС весьма трудоемки, поэтому во многих случаях целесообразно использовать различные приближенные аналитические методы. Один из таких методов базируется на решении дифференциальных уравнений методом Галеркина, который хотя и дает приближенное решение, но позволяет получить любое приближение к точному решению за счет увеличения числа итераций. Важным достоинством метода является то, что повышение порядка исходного уравнения принципиально не усложняет решения. Следует, однако, подчеркнуть, что и приближенные аналитические методы анализа автоколебаний ЭС весьма трудоемки.

Поскольку ЭС является типичной нелинейной системой, к ее исследованию полностью применим метод гармонического баланса, рассмотренный в гл. 3. Метод применим лишь к тем ЭС, которые обладают свойством фильтра низких частот.

5.6. УСТОЙЧИВОСТЬ ЭКСТРЕМАЛЬНЫХ СИСТЕМ. УЛУЧШЕНИЕ КАЧЕСТВА УПРАВЛЕНИЯ В ЭКСТРЕМАЛЬНЫХ СИСТЕМАХ

Теория систем экстремального регулирования (СЭР) была разработана в основном в 60-х годах. Большой вклад в развитие теории СЭР внесли работы В. В. Казакевича, которые использованы при написании параграфов 5.6 и 5.7.

Устойчивость экстремальных систем определяется устойчивостью автоколебательных движений в них. В неустойчивой ЭС

автоколебания принимают расходящийся характер, система удаляется от экстремума и становится неработоспособной.

Условия устойчивости ЭС рассмотрим с позиций классической теории устойчивости А. М. Ляпунова, в соответствии с которой в устойчивой динамической системе достаточно малый сдвиг начальных условий приводит к возникновению возмущенного движения, сколь угодно мало отличающегося от невозмущенного.

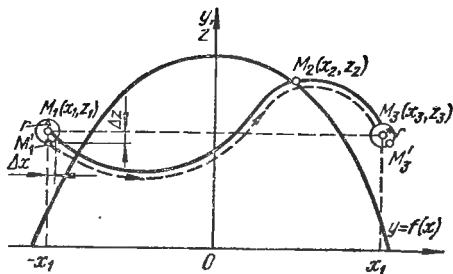


Рис. 5.10. Устойчивость экстремальной системы

Допустим, что периодическое движение в ЭС началось из состояния, характеризуемого точкой $M_1(x_1, z_1)$ (рис. 5.10), объект имеет симметричную статическую характеристику $y=f(x)$, а решение дифференциального уравнения системы имеет вид $z=\varphi(x, x_1, z_1)$. Допустим, что реверс исполнительного механизма произошел в точке фазовой плоскости $M_3(x_3, z_3)$. В этом случае можно записать

$$z_3 = \varphi_2(x_1, z_1) - z_H; \quad x_3 = \varphi_3(x_1, z_3, z_1) = \varphi_3[\varphi_2(x_1, z_1), x_1, z_1].$$

Предположим, что возмущения, приложенные к системе, нарушили периодическое движение и следующий реверс осуществится не в точке $M_1(x_1, z_1)$, а в точке M_1' , отстоящей от точки M_1 на расстоянии r .

Охватим точки M_1 и M_3 окружностями радиуса $r = \sqrt{(\Delta x)^2 + (\Delta z)^2}$. Положим, что следующий цикл периодического движения начинается в точке M_1' или любой другой точке, находящейся на окружности радиуса r , проведенной из точки M_1 как из центра. Движение системы будет устойчивым по теории Ляпунова, если следующий реверс исполнительного механизма произойдет в некоторой точке M_3' с координатами, лежащими внутри окружности радиуса r , проведенного из точки M_3 как из центра, причем радиус r достаточно мал.

Дальнейшие рассуждения справедливы для условий, при которых функции z_3 и x_3 имеют непрерывные частные производные по всем переменным в любой точке внутри окружностей радиуса r и с центрами в точках M_1 и M_3 . Разложим функции z_3 и x_3 в ряды Тейлора, приняв приращения Δz и Δx столь малыми, что можно пренебречь членами разложения выше первого порядка. После разложения при таких условиях получим

$$\begin{aligned} z_3(z_1 + \Delta z, x_1 + \Delta x) &= \\ &= \varphi_2(x_1, z_1) + \Delta x \frac{\partial \varphi_2(x_1, z_1)}{\partial x_1} + \Delta z \frac{\partial \varphi_2(x_1, z_1)}{\partial z_1} - z_H; \end{aligned} \quad (5.32)$$

$$\begin{aligned} x_3(z_1 + \Delta z; x_1 + \Delta x) &= \varphi_3[\varphi_2(x_1, z_1), x_1, z_1] + \\ &+ \Delta x \frac{\partial \varphi_3[\varphi_2(x_1, z_1), x_1, z_1]}{\partial x_1} + \Delta z \frac{\partial \varphi_3[\varphi_2(x_1, z_1), x_1, z_1]}{\partial z_1}. \end{aligned} \quad (5.33)$$

Перепишем (5.32) и (5.33) в сокращенной форме, учитывая, что φ_3 — сложная функция, зависящая не только от начальных условий, но и от функции φ_2 , которая также зависит от начальных условий:

$$z_3 = \varphi_3 + \Delta x \frac{\partial \varphi_2}{\partial x_1} + \Delta z \frac{\partial \varphi_2}{\partial z_1} - z_H; \quad (5.34)$$

$$\begin{aligned} x_3 = & \varphi_3 + \Delta x \frac{\partial \varphi_3}{\partial x_1} + \Delta z \frac{\partial \varphi_3}{\partial z_1} + \\ & + \Delta x \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} + \Delta z \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1}; \end{aligned} \quad (5.35)$$

Так как в процессе автоколебаний фазовая траектория $z(x)$ замкнута, то $z_3 = z_1$; $x_3 = -x_1$, т. е.

$$\begin{cases} \varphi_2(x_1, z_1) - z_H = z_1; \\ \varphi_3[\varphi_2(x_1, z_1), x_1, z_1] = -x_1. \end{cases} \quad (5.36)$$

Преобразуем (5.34) и (5.35) с учетом (5.36):

$$z_3 = z_1 + \Delta x \frac{\partial \varphi_2}{\partial x_1} + \Delta z \frac{\partial \varphi_2}{\partial z_1}; \quad (5.37)$$

$$\begin{aligned} x_3 = & -x_1 + \Delta x \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right) + \\ & + \Delta z \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right); \end{aligned} \quad (5.38)$$

Возмущения вызовут асимметрию автоколебаний. Степень асимметрии проявляется в изменениях координат, очевидных из (5.37) и (5.38):

$$\Delta z = z_3 - z_1 = \Delta x \frac{\partial \varphi_2}{\partial x_1} + \Delta z \frac{\partial \varphi_2}{\partial z_1}; \quad (5.39)$$

$$\begin{aligned} |\Delta x| = |x_3| - |x_1| = & \Delta x \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right) + \\ & + \Delta z \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right); \end{aligned} \quad (5.40)$$

Условие устойчивости автоколебаний системы по Ляпунову, требующее попадания изображающей точки M_3 возмущенного движения внутрь окружности радиуса r , формализуется в виде

$$\begin{aligned} (\Delta z^2) + (\Delta x^2) > & \left[\Delta x \frac{\partial \varphi_2}{\partial x_1} + \Delta z \frac{\partial \varphi_2}{\partial z_1} \right]^2 + \\ & + \left[\Delta x \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right) + \Delta z \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right) \right]^2 \end{aligned}$$

или

$$(\Delta \dot{x}^2) \left[1 - \left(\frac{\partial \varphi_2}{\partial x_1} \right)^2 - \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right)^2 \right] -$$

$$\begin{aligned}
& -2\Delta x \Delta z \left[\frac{\partial \varphi_2}{\partial x_1} \frac{\partial \varphi_2}{\partial z_1} + \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right) \times \right. \\
& \quad \left. \times \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right) \right] + \\
& + (\Delta z)^2 \left[1 - \left(\frac{\partial \varphi_3}{\partial z_1} \right)^2 - \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right)^2 \right] > 0.
\end{aligned}$$

Последнее выражение приведем к квадратичной форме вида

$$D_1 = a_1 (\Delta x)^2 - 2a_3 \Delta x \Delta z + a_2 (\Delta z)^2 > 0, \quad (5.41)$$

где

$$\left\{ \begin{aligned} a_1 &= 1 - \left(\frac{\partial \varphi_2}{\partial x_1} \right)^2 - \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right)^2; \\ a_2 &= 1 - \left(\frac{\partial \varphi_2}{\partial z_1} \right)^2 - \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right)^2; \\ a_3 &= \frac{\partial \varphi_2}{\partial x_1} \frac{\partial \varphi_2}{\partial z_1} + \left(\frac{\partial \varphi_3}{\partial x_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial x_1} \right) \times \\ & \quad \times \left(\frac{\partial \varphi_3}{\partial z_1} + \frac{\partial \varphi_3}{\partial z_3} \frac{\partial z_3}{\partial z_1} \right) \end{aligned} \right\} \quad (5.42)$$

Условие устойчивости (5.41) соблюдается при

$$\begin{vmatrix} a_1 & -a_3 \\ -a_3 & a_2 \end{vmatrix} > 0; \quad a_1 > 0. \quad (5.43)$$

Неравенство (5.43) является критерием устойчивости автоколебаний в системе экстремального управления.

Как видно из изложенного, проверка ЭС на устойчивость связана с определением частных производных координат реверса по всем переменным.

Наиболее тяжелым режимом работы ЭС с точки зрения устойчивости является отработка монотонных возмущений, вызывающих дрейф статической характеристики объекта управления. Такие возмущения могут привести систему в состояние, соответствующее одному из крайних положений исполнительного механизма. Работоспособность таких ЭС обеспечивается введением в систему специального устройства — коммутатора, предназначенного для организации принудительных проверочных реверсов, что обеспечивает немонопольное движение системы к экстремуму.

Если экстремальная система не отвечает требуемым показателям качества: неустойчива или не обеспечивает заданных параметров автоколебаний, имеет недопустимо большие потери на поиск, малое быстродействие, то необходимо принять меры по улучшению качества управления ЭС. Остановимся на одном из методов повышения качества ЭС, предназначенных для увеличения быстродействия системы, т. е. для сокращения времени переходного процесса без снижения качества в статических режимах работы.

В ЭС время переходного процесса определяется иначе, чем в непрерывных системах: под временем переходного процесса в ЭС

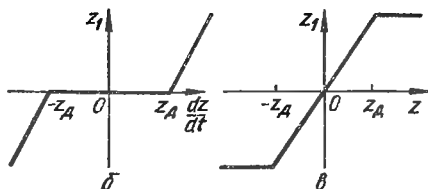
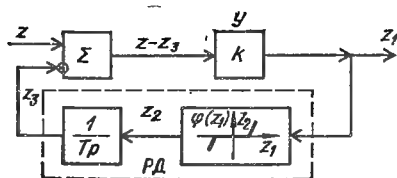


Рис. 5.11. Устройство формирования входного сигнала экстремального регулятора:

а — структурная схема; б, в — статические характеристики

понимается время, отсчитываемое от момента начала поиска экстремума до возникновения колебаний вокруг него.

Обратимся к экстремальной системе (рис. 5.8). Из уравнения (5.29) и рис. 5.8 видно, что для нахождения экстремума функции $f(x)$ при $x = x_{\text{опт}}$, т. е. для выполнения задачи управления, на экстремальный регулятор необходимо подавать функцию $\Phi = T \frac{dz}{dt} + z$.

При большой скорости перемещения исполнительного механизма, т. е. при больших значениях $\frac{dz}{dt}$, параметр z изменится мало, так как время переходного процесса мало. При этих условиях экстремальный регулятор будет отыскивать экстремум функции $y = f(x)$, если на его вход подается только значение $\frac{dz}{dt}$. При этом с увеличением скорости исполнительного механизма точность поиска экстремума функции $y = f(x)$ повышается.

Для измерения производной выходной величины по времени, т. е. для измерения $\frac{dz}{dt}$, в ЭС применяются автоматические компенсаторы (рис. 5.11) — устройства формирования входного сигнала (УФС) экстремального регулятора. Физическая схема УФС включает в себя усилитель (У) с коэффициентом усиления k и отрицательной обратной связью, содержащей интегрирующее звено. Обратная связь реализуется реверсивным конденсаторным двигателем (РД), имеющим зону нечувствительности z_d . Структурная схема РД представляется в виде последовательного соединения интегрирующего звена и нелинейности со статической характеристикой $z_2 = \varphi(z_1)$. В пределах линейных участков статической характеристики передаточная функция УФС, как видно из рис. 5.11, имеет вид

$$W(p) = \frac{k}{1 + k/(Tp)} = \frac{Tp}{1 + Tp/k}$$

В пределах изменения $-z_d < z < z_d$, т. е. в пределах зоны нечувствительности двигателя, $z_1 = kz$, так как обратная связь равна нулю. В результате при медленном изменении z на сигнум-реле

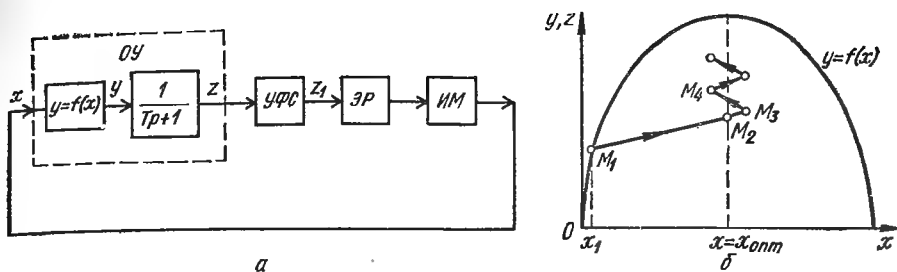


Рис. 5.12. Экстремальная система с устройством формирования входного сигнала регулятора:

а — функциональная схема; б — движение системы к экстремуму

поступает сигнал, пропорциональный выходной величине объекта z , и поиск экстремума осуществляется по параметру z .

Использование УФС обеспечивает поиск экстремума статической характеристики объекта управления с высокой скоростью без перехода на противоположную ветвь характеристики $y=f(x)$. Это значительно уменьшает время поиска экстремума. Функциональная схема такой системы приведена на рис. 5.12, а. Рассмотрим характер движения к экстремуму в этой системе.

Допустим, что исполнительный механизм (ИМ) включился на возрастание входной величины x объекта управления (ОУ) при $x=x_1$, т. е. в точке M_1 статической характеристики $y=f(x)$ (рис. 5.12, б). В точке M_2 при $x=x_{\text{опт}}$ производная выходной величины $\frac{dz}{dt}$ достигает максимального значения. Последующее увеличение x приведет систему в состояние, характеризующееся точкой M_3 , в которой производная станет отличаться от своего максимального значения на величину зоны нечувствительности сигнум-реле. Это вызовет реверс ИМ и уменьшение x до некоторой точки M_4 , в которой вновь произойдет реверс ИМ. В дальнейшем характер движения повторяется и система будет реверсироваться в районе $x=x_{\text{опт}}$. При этом $\frac{dz}{dt} \approx \left(\frac{dz}{dt} \right)_{\text{max}}$ и выходная величина z объекта управления со скоростью, близкой к максимальной, стремится к экстремуму.

Сокращение времени поиска экстремума без УФС, а только за счет увеличения скорости исполнительного механизма неэффективно, так как увеличивает амплитуду автоколебаний и потери на поиск. С другой стороны, поиск экстремума по величине $\frac{dz}{dt}$ может оказаться затруднительным в случаях плохо выраженных экстремумов статических характеристик объектов ЭС; в этих случаях сложно реализовать устройство, измеряющее величину $\frac{dz}{dt}$.

5.7. САМОНАСТРАИВАЮЩИЕСЯ И СЕООБУЧАЮЩИЕСЯ ДИНАМИЧЕСКИЕ СИСТЕМЫ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Кибернетика — наука об общих законах получения, хранения, передачи и преобразования информации в сложных управляющих системах различной физической природы определена Н. Винером как наука об управлении и связи в животном и машине. Развитие кибернетики, появление и развитие новых технических средств радиоэлектроники, вычислительной техники привело к созданию и внедрению самонастраивающихся и самообучающихся систем автоматического управления — наиболее прогрессивных систем автоматики.

Самонастраивающиеся системы (СНС) — это системы, в которых в процессе функционирования автоматически изменяются некоторые параметры управляющей части или управляющей части и объекта управления с целью обеспечения заданного качества управления в условиях нестационарности объекта управления, задающих и возмущающих воздействий.

Самообучающиеся (самоприспосабливающиеся) системы (СОС) — это системы, алгоритм функционирования которых вырабатывается и совершенствуется в процессе самообучения под влиянием внешних возмущений.

В зависимости от условий функционирования предусматривается изменение каких-либо параметров элементов СНС, ее структуры, введение логических элементов, вычислительных устройств, элементов автоматического поиска. В СНС управляющие воздействия формируются не только в функции ее координат, но и в зависимости от состояния внешней среды, проявляющегося в изменениях возмущающих воздействий на систему. Элементы самонастройки призваны существенно повысить качество управления в условиях трудно прогнозируемых или вообще непрогнозируемых возмущений. Вместе с тем введение самонастройки существенно удорожает систему, усложняет ее структуру, снижает показатели надежности.

Функциональная схема СНС представлена на рис. 5.13, а. Она состоит из основного контура, образованного объектом управления (ОУ) и регулятором (Р), и контура самонастройки, включающего

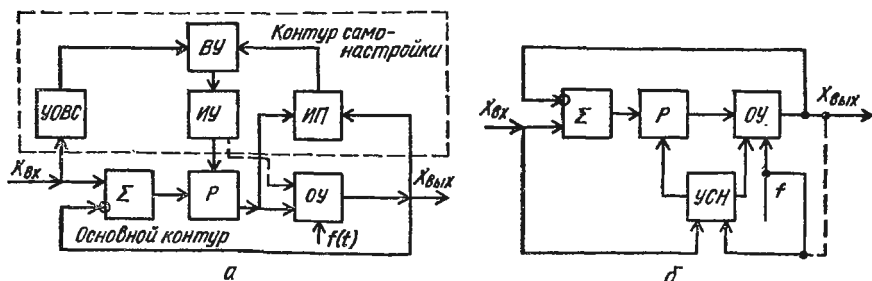


Рис. 5.13. Функциональные схемы самонастраивающейся системы управления:
а — с контуром самонастройки; б — с настройкой по внешним воздействиям

в себя идентификатор процесса (ИП), устройство оценки входных сигналов (УОВС), вычислительное устройство (ВУ) и исполнительное устройство (ИУ). Идентификатор процесса и УОВС составляют контролируемую часть, в задачу которой входит выработка текущей информации о состоянии объекта и внешней среды.

ИП предназначен для определения динамических свойств объекта управления. При этом наиболее полная оценка ОУ заключается в определении его переходных и частотных характеристик, уравнений, передаточной функции и т. п. В более простых случаях в задачу ИП входит формирование отдельных показателей динамики ОУ в дискретной форме.

УОВС предназначено для анализа характера внешних возмущений.

ВУ вырабатывает или хранит в памяти условие самонастройки, которое в наиболее совершенных СНС является критерием оптимального управления объектом. В наиболее сложных системах управления в качестве вычислительного устройства используются электронно-вычислительные машины, как цифровые, так и аналоговые.

ВУ на основе информации о состоянии объекта и внешней среды, т. е. на основе информации, поступающей от ИП и УОВС, вырабатывает требования к характеристикам регулятора. Информация с выхода вычислительного устройства с помощью исполнительного устройства ИУ преобразуется в воздействие на регулятор, обеспечивая требуемый в данный момент закон управления объектом.

Для реализации самонастройки регулятор СНС, а иногда и объект имеют изменяемую часть, характеризующуюся изменяемой структурой и параметрами.

Наиболее характерной чертой СНС является режим автоматического поиска, организуемый с целью выявления отклонений системы от оптимального состояния. В частности, такой поиск осуществляется организацией автоматических пробных движений СНС с анализом входных и выходных переменных.

В тех случаях, когда закон изменения характеристик объекта для различных моментов времени может быть заранее определен, можно применять программную настройку, упрощающую структуру СНС. Однако в общем случае закон изменения характеристик неизвестен и требуется вводить элементы самонастройки.

По исходному принципу самонастройки СНС подразделяются на системы с самонастройкой по сигналам внешних воздействий; с самонастройкой по динамическим характеристикам объектов; комбинированные, использующие для самонастройки сигналы внешних воздействий и информацию о динамических характеристиках объектов.

По способу воздействия элементов самонастройки на систему СНС подразделяются на системы с автоматической настройкой параметров и системы с автоматической настройкой структуры. Последний тип систем получил название *самоорганизующихся*.

По контуру самонастройки и по контуру основной обратной связи СНС могут быть разомкнутыми и замкнутыми.

Заканчивая общую характеристику самонастраивающихся автоматических систем, отметим их наиболее характерные черты: наличие двух контуров — основного и контура самонастройки; включение элементов, способных в процессе работы изменять алгоритм функционирования, параметры, характеристики или структуру, причем эти изменения могут быть как непрерывными, так и дискретными; использование вычислительных устройств для выработки и хранения условий самонастройки (критерия оптимизации); высокая чувствительность к изменению параметров возмущающих и управляющих воздействий; автоматический поиск оптимального состояния при случайных воздействиях на систему.

На рис. 5.13, б показана функциональная схема СНС с настройкой по сигналам внешних воздействий. Система содержит устройство самонастройки (УСН), обеспечивающее приведение системы в режим, оптимальный по выбранному критерию, который в общем случае может быть переменным. УСН анализирует входной сигнал $X_{вх}$, возмущения и помехи f и вырабатывает воздействие, под влиянием которого изменяются параметры или структура регулятора и объекта. По контуру самонастройки система является разомкнутой, если осуществляются измерение и анализ внешних воздействий. Если такого измерения осуществить нельзя, то влияние возмущений анализируется по выходному сигналу $X_{вых}$. В этом случае система будет замкнутой как по основному контуру, так и по контуру самонастройки (см. штриховую связь УСН с выходной величиной $X_{вых}$).

В тех случаях, когда возмущения изменяют параметры объекта и его динамические характеристики, самонастройка осуществляется по динамическим характеристикам объектов управления. Для этой цели широко используются динамические модели, и самонастройка осуществляется по характеристикам модели.

СНС с моделями управляемых объектов может строиться с вычислителем параметров объектов, с эталонной или подстраиваемой моделью, с анализатором характеристик объекта и по другим, в том числе комбинированным принципам.

В СНС с моделью и вычислителем параметров объекта управления текущие значения параметров объекта сравниваются с параметрами моделей. Такие системы имеют относительно сложный контур самонастройки и низкую ее скорость, обусловленную длительностью вычислений параметров объекта. В связи с этим СНС подобного типа эффективны лишь в случаях, когда параметры объекта изменяются медленно по сравнению с длительностью переходных процессов в системе.

В самонастраивающихся системах с эталонной моделью (рис. 5.14) характеристики объекта управления (ОУ) сравниваются с характеристиками эталонной модели (ЭМ). Вектор состояния системы ΔX определяется с помощью идентификатора процесса управления (ИП), выходной вектор которого в устройстве сравнения Σ сравнивается с вектором состояния x_m эталонной модели. Сиг-

нал рассогласования ΔX поступает на устройство самонастройки (УСН), которое формирует воздействия, изменяющие параметры или структуру объекта управления (ОУ), регулятора (Р), главной обратной связи (ОС). В зависимости от сложности системы УСН может осуществлять настройку по всем трем указанным каналам или по одному из них. Во всех случаях задача контура самонастройки сводится к приведению системы к оптимальному состоянию по требуемому критерию функционирования.

В комбинированных самонастраивающихся системах используются оба рассмотренных выше принципа: самонастройка по сигналам внешних воздействий и самонастройка по динамическим характеристикам объекта. Комбинированная СНС также может быть построена на основе эталонной модели, но в этом случае динамические характеристики ЭМ не остаются стабильными, а автоматически изменяются в соответствии с изменениями возмущений f и входных воздействий $X_{вх}$. Для реализации комбинированного принципа самонастройки эталонная модель снабжается дополнительным контуром самонастройки по переменным $X_{вх}$ и f .

Вместо эталонных моделей в СНС для самонастройки могут использоваться анализаторы характеристик объекта и вычислительные устройства. В этом случае определяемые динамические характеристики системы сравниваются с заданными характеристиками. В качестве динамических характеристик, используемых для оценки качества управления в замкнутых системах автоматического управления, могут быть переходные характеристики, импульсные переходные характеристики, частотные функции. С этой целью в СНС предусматриваются средства создания режимов, необходимых для получения указанных характеристик.

Решение задачи синтеза контура самонастройки включает следующие основные этапы: анализ причин, обосновывающих необходимость самонастройки; выбор критерия самонастройки и принципа построения контура самонастройки; разработку алгоритмов идентификации объекта; определение закона изменения настраиваемых параметров; разработку эталонной модели или анализатора характеристик.

Задача синтеза контура самонастройки СНС характеризуется большой сложностью, обусловленной нелинейностью объекта с переменными случайными параметрами, связанными сложными зависимостями с возмущающими воздействиями. Вся задача решается в два укрупненных этапа — первичная оптимизация, когда разрабатывается основной контур, и вторичная оптимизация, связанная с разработкой контура самонастройки. Первичная оптимизация решается с использованием ранее рассмотренных методов синтеза оптимальных динамических систем.

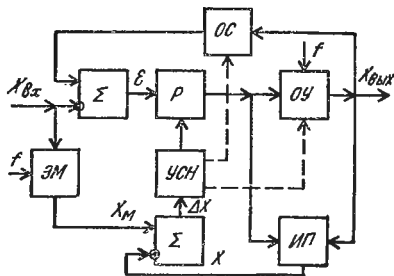


Рис. 5.14. Самонастраивающаяся система с эталонной моделью

В самообучающихся системах (СОС) параметры, структура, закон управления изменяются по сложным алгоритмам, которые совершенствуются на базе результатов самонастройки.

Разработка и эксплуатация самообучающихся систем тесно связаны с проблемой распознавания образов. Системы, построенные на базе распознавания образов, получили название *персептрон* (от англ. *to percept* — воспринимать).

Персептрон является обучаемой распознающей системой, реализующей корректируемое линейное решающее правило в пространстве фиксированных, случайно выбранных признаков входных сигналов. Под решающим правилом понимается алгоритм, позволяющий по результатам измерений определенных признаков объекта (ситуации) принять решение о значениях его параметров, непосредственно не наблюдаемых при измерениях. Решением может быть, например, отнесение интересующего объекта к тому или иному классу объектов по установленной классификации.

Принцип распознавания образов заключается в том, чтобы построить поверхности, разделяющие гиперпространство на конечное число областей, каждая из которых имеет свой образ. Способность системы к «распознаванию» означает, что известные типы объектов будут приписываться к определенному классу, а если появляется неизвестный тип, не относящийся ни к одному известному классу, то ему приписывается новое подразделение гиперповерхности. Распознавание образов после обучения состоит в проверке новых типов, о принадлежности которых к определенному классу ничего не известно. Самообучающиеся системы функционируют, как правило, на базе ЭВМ.

Обеспечение оптимального управления при неполной информации об объекте управления связано с понятием *дуального управления*, т. е. такого, в котором управляющие воздействия имеют двойственный характер — используются для изучения объекта управления и для приведения его в требуемое состояние. При этом процессы изучения объекта управления и процессы управления им тесно связаны и образуют единый сложный двойственный, или дуальный, процесс, определяющий качество функционирования системы управления. В общем случае для отыскания оптимальной стратегии дуального управления используются методы динамического программирования.

Методы решения задач дуального управления связаны со значительными трудностями, поэтому на практике часто ограничиваются отысканием субоптимальных стратегий путем упрощения постановки задачи или сужения класса допустимых стратегий. В частности, для некоторых типов объектов оптимальной может оказаться стратегия, направленная в каждый момент времени только на приведение объекта к требуемому состоянию и не содержащая в себе специальных функций по изучению объекта.

ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ В ТЕОРИИ И ПРАКТИКЕ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

6.1. ТИПЫ ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНЫХ МАШИН

Многие задачи, решаемые с помощью современных автоматических систем, требуют формирования таких сложных законов управления объектами, которые не могут быть реализованы традиционными элементами и устройствами автоматики. Так, например, в системах управления движущимися объектами требуются сложные вычисления, связанные с операциями преобразования координат, решением прямоугольных и сферических треугольников, считением пути и т. п. Очень сложные вычисления производятся в адаптивных системах управления. Анализ и синтез сложных динамических систем также связан с необходимостью решения сложных математических задач. Эти задачи решаются с помощью современных средств вычислительной техники, вводимых в контур управления динамической системой или используемых для разнообразных расчетов и поисков оптимальных решений при анализе и синтезе систем управления.

В современной теории и практике управления динамическими системами используются электронно-вычислительные машины (ЭВМ) различных типов, отличающиеся принципом действия, составом элементной базы, возможностями использования в технических системах управления.

По принципу действия все электронно-вычислительные машины, используемые в динамических системах управления, разделяются на два типа: 1) аналоговые (АВМ); 2) цифровые (ЭЦВМ или, чаще, ЭВМ). Разные принципы действия, разная материальная база указанных типов ЭВМ обусловили разные сферы их применения в теории и практике управления динамическими системами.

Аналоговые электронно-вычислительные машины представляют собой вычислительные устройства непрерывного (аналогового) типа. Результат вычисления получается в виде непрерывного электрического сигнала, отражающего значение определяемой переменной, выводимого на экран осциллографа. Выходной сигнал АВМ может быть также использован как управляющее воздействие в автоматической системе, а сама АВМ может сопрягаться с регулятором в качестве его составной части.

Принцип действия АВМ и ее конструкция просты, удобны, быстро осваиваются техническим персоналом. Машина легко сопрягается с элементами систем непрерывного управления. В качестве вычислительной системы АВМ наиболее удобна и эффективна для решения дифференциальных уравнений как линейных, так и нелинейных, особенно с типовыми нелинейностями. Для этого АВМ снабжена набором функциональных блоков, модулирующих нели-

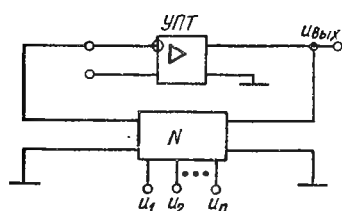


Рис. 6.1. Схема операционного усилителя

нейности. Порядок дифференциальных уравнений, решаемых на АВМ, различен для различных типов АВМ, однако этот вопрос не имеет принципиального значения, так как аналоговая вычислительная система легко наращивается из отдельных блоков и машин в целом.

Основу всякой АВМ составляет операционный усилитель (ОУ) — усилитель постоянного тока (УПТ) с глубокой отрицательной обратной связью и многополюсником N (рис. 6.1).

В состав многополюсника входят резисторы, конденсаторы, делители напряжения, диодные ограничители и другие нелинейные узлы, содержащие элементы с естественными нелинейными характеристиками, например варисторы — полупроводниковые резисторы с нелинейно изменяющимся электрическим сопротивлением. Различные схемы многополюсников совместно с УПТ обеспечивают ту или иную функциональную зависимость между входными и выходными электрическими напряжениями, т. е. обеспечивают выполнение той или иной математической операции над реальными переменными, которые замещаются электрическими напряжениями.

В простейшем случае операционный усилитель совершает инвертирование входного сигнала (рис. 6.2, а). Чтобы убедиться в этом, примем во внимание, что УПТ имеет высокий коэффициент усиления и очень малый входной ток. Поэтому точка суммирования токов i_1 и i_0 является практически потенциально заземленной и ее потенциал близок к нулю, т. е. $u_{\Sigma} \approx 0$, $u_{вх} \approx 0$. С учетом этого для схемы (рис. 6.2, а) можно записать

$$i_1 = u_1/z_1; \quad i_0 = u_{\text{вых}}/z_0; \quad i_1 + i_0 = 0,$$

откуда

$$\frac{u_{\text{вых}}}{u_1} = -\frac{z_0}{z_1}.$$

Переходя к операторной форме, получим

$$\frac{u_{\text{вых}}(p)}{u_1(p)} = -\frac{z_0(p)}{z_1(p)} = -W(p). \quad (6.1)$$

Из выражения (6.1) видно, что передаточная функция линейного операционного усилителя целиком определяется параметрами входной цепи и цепи обратной связи. Таким образом, изменяя ха-

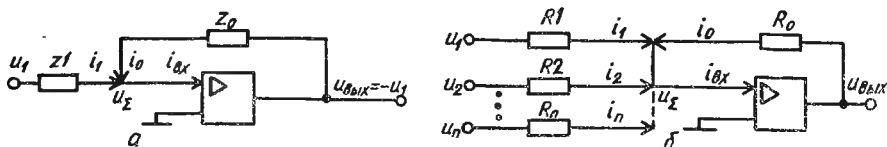


Рис. 6.2. Схемы реализации математических операций:

а — инвертирования; б — суммирования

раक्टर соотношения z_0/z_1 , можно обеспечивать различные функциональные связи между входными и выходными напряжениями операционного усилителя, а следовательно, между теми переменными, которые замещаются этими напряжениями.

В зависимости от требуемой математической операции во входную цепь и цепь обратной связи включаются высокостабильные резисторы, конденсаторы, делители напряжения. В результате формируются аналоговые решающие блоки, осуществляющие разнообразные математические операции — сложение, умножение и деление, дифференцирование, интегрирование электрических напряжений, связанных известными соотношениями с исследуемыми переменными состояниями динамической системы управления.

Например, для схемы (рис. 6.2, б) имеем:

$$i_l = u_l/R_l, \quad l = \overline{1, n}; \quad i_0 = u_{\text{вых}}/R_0; \quad \sum_{l=1}^n i_l + i_0 = 0,$$

откуда

$$u_{\text{вых}} = - \sum_{l=1}^n \frac{R_0}{R_l} u_l, \quad l = \overline{1, n}.$$

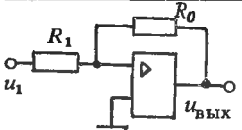
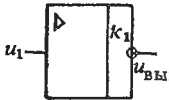
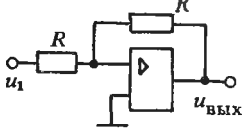
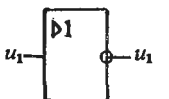
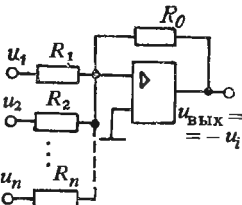
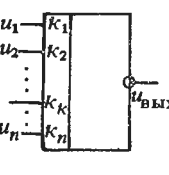
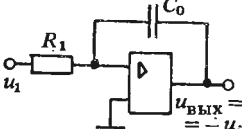
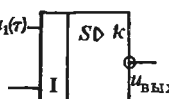
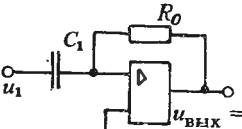
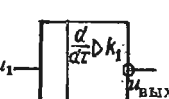
Таким образом, аналоговый решающий блок (рис. 6.2, б) может быть использован для операции суммирования переменных, замещаемых в АВМ электрическими напряжениями. Проведя аналогичные рассуждения, можно установить схемы аналоговых решающих блоков для других математических операций. В табл. 6.1 приведены схемы решающих блоков для выполнения ряда линейных математических операций, их названия и условные обозначения, принятые в практике моделирования динамических систем на АВМ.

Конструктивно АВМ собирается в виде совокупности решающих блоков, организованных в вычислительную систему с помощью электрических связей так, что результат математической операции передается с выхода одного блока на входы других. Исходные исследуемые переменные замещаются физическими величинами — электрическими напряжениями с использованием соответствующих масштабов. Результат решения задачи на АВМ представляется в виде электрических напряжений, которые можно измерять приборами, регистрировать с помощью записывающих приборов, наблюдать на экранах осциллографов.

Обычный состав типовой АВМ включает в себя следующие функциональные части: операционные усилители, наборное поле, устройства управления, измерительную и регистрационную аппаратуру, источники питания.

Операционные усилители обеспечивают выполнение математических операций, часть из которых рассмотрена выше и приведена в табл. 6.1. Кроме линейных блоков в состав АВМ входят блоки функциональных преобразователей — блоки нелинейных функций для воспроизведения типовых нелинейностей. Кроме того, имеются блоки запаздывания, программного управления, релейные блоки и др.

Таблица 6.1

Схема операционного усилителя	Наименование	Уравнение	Условное обозначение
	Инвертирующий с усилением k_1	$u_{\text{ВЫХ}} = -k_1 u_1$ $k_1 = R_0/R_1$	
	Инвертирующий с коэффициентом усиления 1	$u_{\text{ВЫХ}} = -u_1$	
	Суммирующий	$u_{\text{ВЫХ}} = -\sum_{k=1}^n k_k u_k$ $u_i = \sum_{k=1}^n k_k u_k$ $k_k = R_0/R_k; k = 1, n$	
	Интегрирующий	$u_{\text{ВЫХ}} = k_1 \left[\int_0^{\tau} u_1(\tau) d\tau + u_{\text{ВЫХ}}(0) \right]$ $k_1 = 1/R_1 C_0$	
	Дифференцирующий	$u_{\text{ВЫХ}} = -k_1 \frac{du_1}{d\tau}$ $u_i = k_1 \frac{du_1}{d\tau}; k_1 = C_1 R_0$	

В режиме подготовки устанавливаются необходимые характеристики и параметры операционных усилителей. Решение задачи может быть как однократное, так и многократное, автоматически повторяемое. Автоматическое прерывание решения может использоваться для фиксирования и измерения машинных переменных — электрических напряжений — или для изменения схемы набора на коммутационном поле АВМ. Наконец, режим возврата АВМ в исходное состояние требует при необходимости повторно воспроизвести однократное решение задачи.

Наборное поле (поле коммутации) предназначено для организации соединений между отдельными решающими блоками. Эти соединения осуществляются с помощью электрических проводников — коммутационных шнуров с однополюсными вилками, вставляемыми в гнезда наборного поля. В отдельных, обычно больших, АВМ наборное поле может быть съемным. Это повышает

эффективность их использования, так как позволяет совмещать операции решения одной задачи и подготовки другой.

Устройства управления предназначены для организации взаимодействия во времени отдельных частей АВМ. Устройство управления выполняется в виде панели или пульта управления, с помощью которого оператор настраивает АВМ на требуемый режим работы — подготовку, решение задачи, автоматическое прерывание решения, возврат машины в исходное состояние.

В состав *измерительной и регистрирующей аппаратуры* АВМ входят стрелочные измерительные приборы, в основном вольтметры, цифровые и печатающие приборы. Измерительная аппаратура служит для настройки операционных усилителей при подготовке АВМ к решению задачи и для измерения машинных переменных при фиксировании решений. В состав регистрирующей аппаратуры входят однолучевые и многолучевые осциллографы с длительным послесвечением экрана. Используются также шлейфовые осциллографы для записи результатов решения на бумаге.

Источники питания предназначены для преобразования напряжений промышленной электрической сети в рабочие напряжения АВМ — 10, 25, 50, 100 В и др. Основное требование к источникам питания АВМ — высокая стабильность выходных напряжений, так как изменение рабочих напряжений при решении задач на АВМ воспринимается как изменение машинных переменных и существенно влияет на точность решения.

Имея практически одинаковый набор устройств различных типов, АВМ отличаются количеством операционных блоков, определяющим предельные возможности машины, которые выражаются в основном в порядке дифференциальных уравнений, решаемых на АВМ. По этому признаку АВМ подразделяются на три класса: малые (до 20 операционных блоков), средние (20—60 блоков) и большие (свыше 60 блоков).

Общим недостатком АВМ является низкая точность решения задач, вытекающая из аналогового принципа действия этого типа вычислительных устройств. Возможности АВМ также существенно ограничены из-за отсутствия устройств памяти для хранения больших объемов информации. Широкое внедрение цифровых электронно-вычислительных машин существенно снизило область применения и масштабы использования АВМ. Тем не менее в сфере управления техническими системами и технологическими процессами роль АВМ весьма велика. Эти машины несоизмеримо дешевле и проще, чем ЭЦВМ. Вместе с тем АВМ по отдельным показателям превосходит ЭЦВМ. К таким показателям относятся более высокое быстродействие, вытекающее из того, что АВМ работает в реальном масштабе времени, а также удобство сопряжения с реальными элементами непрерывных автоматических систем. В ряде случаев АВМ эффективно дополняют ЭЦВМ, образуя гибридную вычислительную систему.

Цифровые электронно-вычислительные машины представляют собой мощную высокоэффективную вычислительную систему дискретного типа, широко используемую во всех сферах человеческой

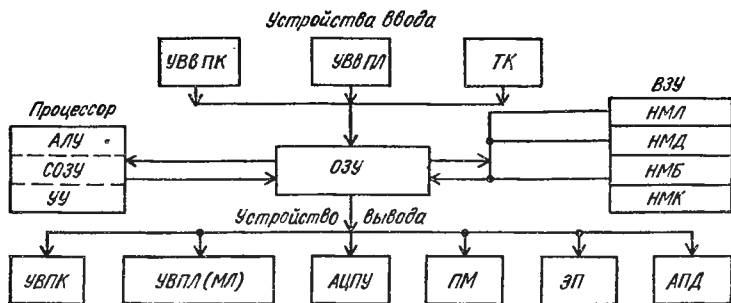


Рис. 6.3. Структура ЭВМ

деятельности. Современная теория и практика управления немислимы без широкого использования ЭВМ при разработке, проектировании и эксплуатации систем управления. Если имеется численный метод решения, то с помощью современных ЭВМ можно решить практически любую задачу в любой области науки, техники, экономики, общественной жизни.

Принципы современных ЭВМ сформировались в 40—50-х годах текущего века. С тех пор сменилось три поколения ЭВМ, каждое из которых значительно превосходило предыдущее по всем решающим показателям, к числу которых относят быстродействие, объем оперативной памяти, возможности операционной системы, массу, габариты, стоимость и ряд других.

В настоящее время в теории и практике управления широко используется множество ЭВМ различных типов, которые подразделяются на три вида: большие ЭВМ, малые, или мини-ЭВМ, микроЭВМ. Несмотря на большое разнообразие типов и конструкций ЭВМ, все они имеют общие принципы, отличаясь характером элементной базы и эффективностью реализации принципов работы. Структура, состав и взаимосвязь основных элементов ЭВМ представлены на рис. 6.3. Эта структура обусловлена содержанием процесса обработки информации на ЭВМ, в состав которого входят следующие основные операции: подготовка исходных данных для ввода в вычислительную машину, ввод исходных данных с машинных носителей информации, решение задачи на ЭВМ, вывод результатов решения.

В соответствии с указанным перечнем функций вычислительной системы ЭВМ включает следующие основные элементы (рис. 6.3): процессор, оперативное запоминающее устройство (ОЗУ), внешние запоминающие устройства (ВЗУ), устройства ввода (УВВ), устройства вывода (УВ).

Процессор — центральное устройство ЭВМ, предназначенное для преобразования информации, заданного программой управления всем вычислительным процессом и взаимодействием устройств вычислительной машины. Основными частями процессора являются арифметико-логическое устройство (АЛУ) и устройство управления (УУ). В быстродействующих ЭВМ в состав процессора вхо-

дит сверхоперативное запоминающее устройство (СОЗУ). Арифметико-логическое устройство осуществляет арифметическое и логическое преобразование информации по командам программы. Устройство управления определяет последовательность выборки команд из памяти, вырабатывает управляющие сигналы, координирует работу ЭВМ, обрабатывает сигналы прерывания программ, осуществляет защиту памяти, контролирует и диагностирует работу процессора. СОЗУ хранит информацию, обрабатываемую процессором и поступающую из ОЗУ.

ОЗУ составляет оперативную память ЭВМ, в которой хранится информация, передаваемая в СОЗУ. При отсутствии СОЗУ информация из ОЗУ в виде команд программы и исходных операндов передается в АЛУ. Из процессора в ОЗУ передаются конечные и промежуточные результаты преобразования информации. В качестве запоминающих элементов в ОЗУ применяются ферритовые кольца в устройствах памяти на магнитных сердечниках или быстродействующие полупроводниковые запоминающие устройства.

ВЗУ — внешняя память ЭВМ, в качестве которой используются накопители на магнитных дисках (НМД), магнитных лентах (НМЛ), магнитных барабанах (НМБ), магнитных картах (НМК).

Два вида памяти — оперативная и внешняя — используются в ЭВМ с тем, чтобы обеспечить два противоречивых с точки зрения физической реализации требования: высокое быстродействие (малое время доступа к хранимой информации) и возможность хранения больших объемов информации. ОЗУ отвечает первому требованию, ВЗУ — второму. СОЗУ обладает быстродействием, примерно на порядок выше, чем ОЗУ. Быстродействующие устройства памяти значительно дороже, чем ВЗУ. Из видов внешней памяти в современных ЭВМ наиболее широко используются НМД и НМЛ. Взаимодействие ОЗУ с процессором и ВЗУ осуществляется устройством управления.

Для подготовки к вводу в ЭВМ информация переносится с документов на машинные носители, которыми являются перфокарты (ПК) и перфоленты (ПЛ). Возможен также ввод информации с телеграфного канала (ТК).

Результаты обработки информации в ЭВМ выводятся на устройства вывода, к которым относятся устройства вывода на перфокарты (УВПК), перфоленту (УВПЛ), магнитную ленту (УВМЛ), алфавитно-цифровое печатающее устройство (АЦПУ), пишущие машинки (ПМ), экранные пульта (ЭП), в аппаратуру передачи данных (АПД). В качестве устройств отображения результатов обработки информации на ЭВМ используются разнообразные устройства знаковой регистрации, например электрохимические, искровые и др., характеризующиеся высокой скоростью печати информации.

Информация в ЭВМ — буквы, цифры, специальные знаки — хранится, как правило, в двоично-кодированном виде, а цифры — в двоичной системе счисления. Это вытекает из природы элементной базы ЭВМ, содержащей быстродействующие элементы с двумя устойчивыми состояниями, отличающимися высокой надеж-

ностью и экономичностью. Двоичная система счисления позволяет также сравнительно просто обеспечить технически выполнение вычислительных операций.

Выполнение программы в ЭВМ — это последовательное осуществление в заданном порядке арифметических и логических операций над словами (кодами), хранящимися в памяти, действий по организации вычислительного процесса и оценка получающихся результатов.

Описанная структура и принцип действия цифровых вычислительных машин в основном являются общими для ЭВМ различных типов. Однако совершенствование элементной базы привело к созданию ЭВМ, существенно отличающихся по своим характеристикам. В технике управления в последние годы все шире используются ЭВМ на основе микропроцессоров — микроЭВМ. *Микропроцессор* — блок полупроводниковой памяти, содержащий множество схемных элементов, монолитно объединенных на очень небольшой площади — в несколько квадратных миллиметров. Степень интеграции микросхем такова, что на одном кристалле кремния может содержаться более 30 000 диодов.

В микроЭВМ микропроцессор используется в качестве блока, осуществляющего информационные связи.

МикроЭВМ (микрокомпьютер), способный решать сложные задачи управления, существенно отличается по своим технико-экономическим данным от предшествующих ЭВМ. Например, первая цифровая вычислительная машина, положившая начало созданию современных ЭВМ, — американский компьютер ЭНИАК — весила 30 т и размещалась в специальном здании со сложной системой вентиляции. Такие же функции, как ЭНИАК, в настоящее время может выполнять микрокомпьютер размером с пачку сигарет, массой до 1 кг, стоимостью в 8 тыс. раз меньше.

МикроЭВМ широко внедряются в технику управления во всех областях народного хозяйства. С их помощью решаются многообразные задачи, связанные со сбором информации о состоянии управляемого процесса, осуществляются контроль за ходом процесса, регулирование, управление и оптимизация управляемых переменных.

Роль микроЭВМ в задачах автоматизации процессов повышается в связи с тем, что управление процессами требует не столько вычислений, сколько логической увязки разнообразной информации. В микроЭВМ значительно легче, чем в универсальных ЭВМ, осуществляется связь с измерительными и исполнительными органами управляемого процесса.

Универсальность цифровых вычислительных машин как средства решения самых разнообразных задач, огромные объемы информации, перерабатываемые и хранимые в ЭВМ, мощные алгоритмические возможности сделали ЭВМ эффективным средством решения современных задач управления. Недостаток цифровых ЭВМ — ограниченное быстродействие и трудность сопряжения с физическими элементами систем автоматического управления.

6.2. ИССЛЕДОВАНИЕ ДИНАМИЧЕСКИХ СИСТЕМ УПРАВЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ АНАЛОГОВЫХ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

В теории и практике управления динамическими системами АВМ используются в двух направлениях: 1) в качестве элементов регуляторов систем автоматического управления; 2) в качестве эффективного средства исследования динамики САУ с целью получения оценки качества управления и синтеза корректирующих устройств.

В первом направлении АВМ позволяют сравнительно легко сформировать и реализовать широкий круг законов управления объектами и процессами, в том числе таких, которые трудно или невозможно сформировать и реализовать на основе традиционных средств автоматики — механических, гидравлических, электрических, пневматических и др.

Во втором направлении АВМ используются как эффективное средство решения дифференциальных уравнений, описывающих динамические процессы в системах автоматического управления.

Если система автоматического управления представлена в виде совокупности типовых структурных звеньев, то АВМ позволяет легко создать аналоговую модель системы путем набора на машине уравнений типовых звеньев.

В качестве примера рассмотрим исследование с помощью АВМ качества управления в системе, структурная схема которой приведена на рис. 6.4. Схема отражает электромашинную систему автоматического управления электроприводом мощного экскаватора.

Разобьем всю совокупность структурных звеньев рассматриваемой автоматической системы на три группы — I, II и III (рис. 6.4). Звенья I группы описываются уравнениями вида

$$\frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{a}{1 + bp}, \quad (6.2)$$

где $x_{\text{вых}}$ и $x_{\text{вх}}$ — выходная и входная величины соответствующего звена.

При исследовании САУ на АВМ реальные физические переменные, характеризующие систему, заменяются машинными переменными — электрическими напряжениями аналоговой вычислительной машины. Натуральное время t также пересчитывается в машинное время τ с использованием масштабного коэффициента времени m_t .

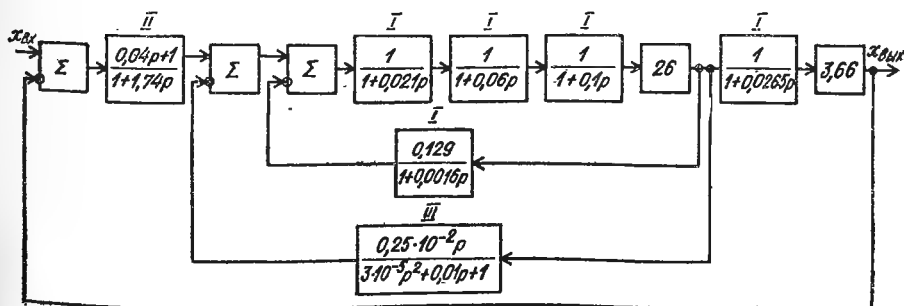


Рис. 6.4. Структурная схема системы автоматического управления

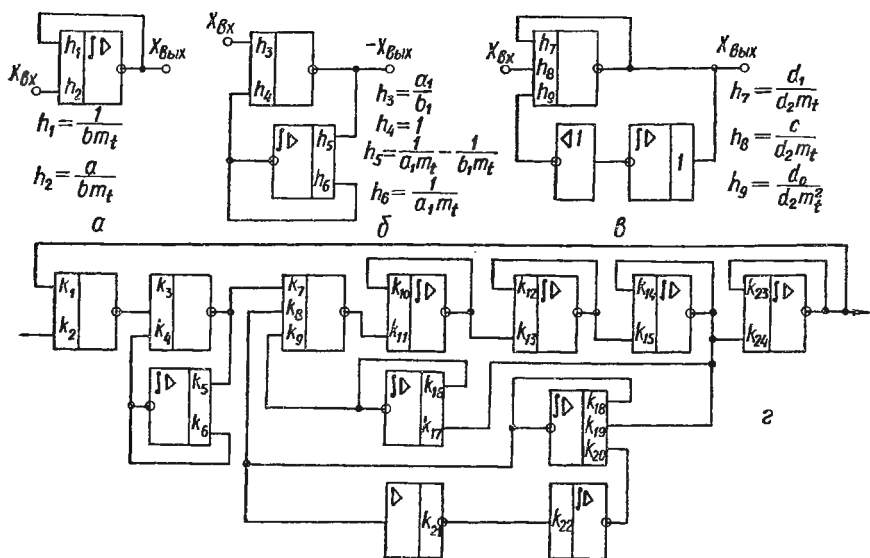


Рис. 6.5. Схемы набора на АВМ моделей:

а — структурных звеньев I группы; *б* — II группы; *в* — III группы; *г* — системы автоматического управления

Учитывая, что $p = \frac{d}{dt}$, в машинных единицах имеем:

$$\tau = m_t t; \quad P = \frac{d}{d\tau} = \frac{p}{m_t}, \quad \text{т. е. } p = m_t P.$$

С учетом сказанного (6.2), выраженное в машинных переменных, принимает вид

$$\frac{X_{\text{ВЫХ}}}{X_{\text{ВХ}}} = \frac{a}{1 + m_t b P}, \quad (6.3)$$

где $X_{\text{ВЫХ}}$ и $X_{\text{ВХ}}$ — выходная и входная машинные величины звена. Преобразуем (6.3) к виду

$$X_{\text{ВЫХ}} P = \frac{a}{b m_t} X_{\text{ВХ}} - \frac{1}{b m_t} X_{\text{ВЫХ}}, \quad (6.4)$$

где a — коэффициент усиления; b — постоянная времени звена.

Уравнение (6.4) является исходным для набора на АВМ аналоговой модели звеньев I группы. Схема набора на АВМ модели структурных звеньев I группы, соответствующая выражению (6.4), приведена на рис. 6.5, *а*.

Чтобы получить исходное уравнение для аналоговой модели звена II группы, приведем уравнение этого звена к виду

$$\frac{x_{\text{ВЫХ}}(p)}{x_{\text{ВХ}}(p)} = \frac{1 + a_1 p}{1 + b_1 p}.$$

Введя машинные переменные, получим

$$X_{\text{ВЫХ}}(1 + m_t b_1 P) = X_{\text{ВХ}}(1 + m_t a_1 P).$$

После некоторых преобразований, найдем

$$X_{\text{ВЫХ}} = \frac{a_1}{b_1} X_{\text{ВХ}} + \frac{1 - a_1/b_1}{1 + m_t a_1 P} X_{\text{ВЫХ}}. \quad (6.5)$$

Уравнение (6.5) является исходным для набора на АВМ модели звена II группы. Сомножитель перед $x_{\text{вых}}$ во втором слагаемом выражения (6.5) отражает модель аperiodического звена с постоянной времени a_1 и коэффициентом усиления $1 - a_1/b_1$. Следовательно, схема набора второго слагаемого правой части (6.5) отличается от схемы набора звена I группы только значением коэффициентов. Роль коэффициента $1/(b m_t)$ в схеме I группы выполняет коэффициент $1/(a_1 m_t)$ в схеме II группы, а коэффициент $a/(b m_t)$ выражается уравнением

$$\frac{1 - a_1/b_1}{a_1 m_t} = \frac{1}{a_1 m_t} - \frac{1}{b_1 m_t}.$$

После набора схемы, соответствующей второму слагаемому выражения (6.5), выходную величину этой схемы нужно сложить с величиной $\frac{a_1}{b_1} x_{\text{вых}}$. Схема набора модели звена II группы приведена на рис. 6.5, б.

Чтобы получить исходное уравнение для аналоговой модели звена III группы, приведем уравнение этого звена к виду

$$\frac{x_{\text{вых}}(p)}{x_{\text{вх}}(p)} = \frac{cp}{d_2 p^2 + d_1 p + d_0},$$

где $c = 0,25 \cdot 10^{-2}$; $d_2 = 3 \cdot 10^{-5}$; $d_1 = 0,01$; $d_0 = 1$.

Введя машинные переменные, получим

$$P X_{\text{вых}} = \frac{c}{d_2 m_t} X_{\text{вх}} - \frac{d_1}{d_2 m_t} X_{\text{вых}} - \frac{d_0}{d_2 m_t^2} \frac{1}{P} X_{\text{вых}}. \quad (6.6)$$

Соответствующая уравнению (6.6) схема аналоговой модели звена III группы приведена на рис. 6.5, в.

Схема набора на АВМ структурной схемы системы автоматического управления, представленной на рис. 6.4, показана на рис. 6.5, г. Эта схема составлена на основании схем набора отдельных звеньев (рис. 6.5, а, б, в) с учетом того, что после прохождения интегрирующего усилителя в АВМ машинная переменная изменяет свой знак на противоположный. В связи с этим в соответствующие участки включены усилители-инверторы.

Аналоговая модель САУ дает возможность легко и наглядно получить картину переходного процесса в системе при заданном возмущении, проследить влияние параметров системы на качество управления, выбрать параметры, обеспечивающие оптимальное по заданному показателю качество управления.

Рассмотрим работу САУ (рис. 6.4) при внешнем воздействии в форме единичного скачка на входе.

Исходя из ориентировочных показателей качества переходного процесса, которые, например, могут быть получены по виду логарифмических частотных характеристик САУ, выбираем масштабный коэффициент времени $m_t = 40$. Тогда коэффициенты $k_1, \dots, k_{24}$ в схеме модели (рис. 6.5, г), определенные по параметрам структурной схемы (рис. 6.4), принимают следующие значения:

$$\begin{aligned} k_1 &= 1; \quad k_2 = 1; \quad k_3 = \frac{0,04}{2,74} = 0,023; \quad k_4 = \frac{1}{40 \cdot 0,04} = 0,625; \\ k_5 &= 0,625 - \frac{1}{40 \cdot 1,74} = 0,611; \quad k_6 = k_4 = 0,625; \quad k_7 = 1; \quad k_8 = 1; \quad k_9 = 1; \\ k_{10} &= \frac{1}{40 \cdot 0,021} = 1,19; \quad k_{11} = k_{10} = 1,19; \quad k_{12} = \frac{1}{40 \cdot 0,06} = 0,416; \\ k_{13} &= k_{12} = 0,416; \quad k_{14} = \frac{1}{40 \cdot 0,1} = 0,25; \quad k_{15} = 26,8 k_{14} = 6,7; \\ k_{16} &= \frac{1}{40 \cdot 1,6 \cdot 10^{-3}} = 15,65; \quad k_{17} = 0,129 k_{16} = 2,02; \end{aligned}$$

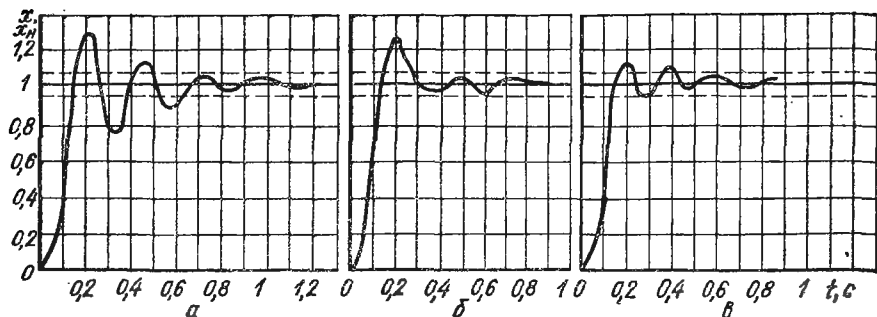


Рис. 6.6. Переходные процессы при моделировании на АВМ системы автоматического управления ($f(t) = [1]$): а — при $T_{г.с} = 0,01$ с; б — $0,02$ с; в — $0,03$ с

$$k_{18} = \frac{0,01}{3 \cdot 10^{-5} \cdot 40} = 8,35; \quad k_{19} = \frac{0,25 \cdot 10^{-2}}{3 \cdot 10^{-5} \cdot 40} = 2,08;$$

$$k_{20} = \frac{1}{3 \cdot 10^{-5} \cdot 40^2} = 20,8; \quad k_{21} = k_{22} = 1; \quad k_{23} = \frac{1}{0,0265 \cdot 40} = 0,945;$$

$$k_{24} = 3,66 \cdot k_{23} = 3,46.$$

При использовании АВМ конкретного типа может оказаться, что не все полученные значения коэффициентов могут быть реализованы на наборном поле. В этом случае значения коэффициентов необходимо преобразовать так, чтобы при условии сохранения общего коэффициента усиления системы получить приемлемые для набора значения отдельных коэффициентов.

Переходный процесс в САУ, снятый с экрана осциллографа АВМ, приведен на рис. 6.6, а, из которого видно, что время переходного процесса составляет $t_{\max} = 0,73$ с, максимальное перегуливание $\sigma_{\max} = 30\%$, колебательность $\mu = 2$.

На рис. 6.6, б, в приведены осциллограммы переходных процессов в САУ (рис. 6.4) при разных постоянных времени цепи гибкой обратной связи. Из приведенных осциллограмм видно существенное влияние постоянной времени цепи гибкой обратной связи на форму переходных процессов. При указанных параметрах лучшую форму имеет переходный процесс при $T_{г.с} = 0,02$ с. В этом случае $t_{\max} = 0,27$ с, $\sigma_{\max} = 23\%$, $\mu = 1$.

Приведенный пример показывает большие возможности АВМ для анализа и синтеза непрерывных систем автоматического управления: переход от одного значения параметров к другому осуществляется простым изменением коэффициентов в схеме набора модели. Модель позволяет без громоздких расчетов просмотреть множество вариантов системы и выбрать оптимальный по заданному показателю вариант.

6.3. ИССЛЕДОВАНИЕ ДИНАМИЧЕСКИХ СИСТЕМ УПРАВЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ ЦИФРОВЫХ ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНЫХ МАШИН

Цифровые электронно-вычислительные машины (в дальнейшем для сокращения будем их называть ЭВМ) являются наиболее эффективным средством исследования систем управления вообще и систем автоматического управления в частности. ЭВМ характери-

зуются универсальностью, высокой точностью результатов, весьма широкими логическими возможностями. Для решения многих задач анализа и синтеза сложных автоматических систем использование ЭВМ особенно эффективно. К этим задачам относится построение областей в пространстве параметров исследуемой системы, внутри которых удовлетворяется критерий качества управления.

Как инструмент решения сложных задач анализа и синтеза САУ ЭВМ используется в двух направлениях.

В первом направлении ЭВМ применяется как эффективное средство реализации сложившихся методов исследования САУ. В настоящее время создано множество программ для ЭВМ по расчету характеристик САУ (например, частотных), устойчивости по алгебраическим и частотным критериям, переходных процессов на основе вещественной частотной характеристики, выделения областей устойчивости в плоскости параметров САУ и др.

Во втором направлении ЭВМ применяется для разработки и реализации новых эффективных методов исследования САУ или для реализации ранее известных, но не используемых методов, связанных с трудоемкими и громоздкими вычислениями. К этому направлению относятся: градиентные и статистические методы оптимизации; линейное и нелинейное программирование; методы динамического программирования; методы оценки качества управления, использующие матричные операции, и др.

Ко второму направлению относится также построение переходных процессов в САУ на основе решения с помощью ЭВМ исходных дифференциальных уравнений исследуемых САУ. Интенсивное развитие высокоэффективных средств вычислительной техники несомненно приведет к резкому возрастанию роли методов исследования качества САУ с помощью ЭВМ путем решения исходных дифференциальных уравнений. Теоретическую основу этих методов составляют численные методы интегрирования дифференциальных уравнений — методы Эйлера, Рунге — Кутта, Адамса и др. Однако в настоящее время реализация этих методов на ЭВМ не всегда оправдана из-за больших затрат машинного времени при анализе САУ, описываемых дифференциальными уравнениями высоких порядков.

Приведенные выше характеристики АВМ и ЭВМ показывают, что весьма перспективным средством анализа и синтеза систем автоматического управления являются вычислительные системы, использующие как аналоговые, так и цифровые электронно-вычислительные машины. Такие системы получили название *гибридных вычислительных систем* (ГВС). Имеющийся опыт использования ГВС показал их высокую эффективность, особенно в тех сложных задачах оптимального управления, в которых требуется проводить большой объем работ по интегрированию дифференциальных уравнений и уравнений в частных производных. Это объясняется тем, что решение дифференциальных уравнений в частных производных на базе цифровых вычислительных машин связано со значительными, часто неразрешимыми, трудностями.

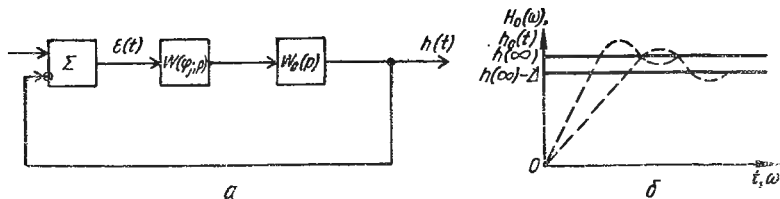


Рис. 6.7. Система управления по интегральной оценке ошибки:
а — структура системы; б — ее динамические характеристики

Объединение АВМ и ЭЦВМ в ГВС и выбор надлежащей структуры ГВС позволяет создавать высокопроизводительные вычислительные системы, обладающие быстродействием во многие десятки миллионов операций в секунду, высокой точностью с развитыми алгоритмическими языками высокого уровня, совершенными средствами отображения результатов обработки информации (графопостроители, дисплеи и т. п.).

В качестве примера исследования САУ на базе ЭВМ рассмотрим задачи синтеза и оптимизации систем управления по интегральной оценке ошибки.

Пусть задана структура системы (рис. 6.7, а), причем часть ее параметров известна, а остальные параметры φ_j , $j = \overline{1, s}$, требуется определить. Неизвестные параметры на схеме локализованы, хотя они могут относиться к нескольким различным элементам системы, как к исходным, так и к корректирующим.

Остановимся на двух задачах — задаче синтеза и задаче оптимизации. Синтез сводится к определению неизвестных параметров системы из условия минимума интегральной оценки приближения реального переходного процесса в системе к желаемому процессу. Оптимизация заключается в определении неизвестных параметров из условия минимума интегральной оценки переходной ошибки. В обоих случаях требуется обеспечить такие обязательные показатели качества управления, как устойчивость, заданные значения перерегулирования, колебательности, статической ошибки.

Поставленные задачи сводятся к минимизации целевой функции, которая представляет собой оценку погрешности воспроизведения желаемого или оптимального процесса фактическим процессом по интегральному квадратичному критерию. В целом поставленная задача как вычислительная сводится к задаче нелинейного программирования.

Изображение переходной функции $h(t)$ выходной величины, по которой определяются неизвестные параметры, представим в виде

$$H(\varphi, p) = \frac{\sum_{i=0}^m b_i(\varphi) p^i}{\sum_{i=0}^m a_i(\varphi) p^i} \doteq h(t), \quad m \leq n, \quad (6.7)$$

где $\varphi = \|\varphi_1, \varphi_2, \dots, \varphi_s\|$ — вектор варьируемых параметров.

В выражении (6.7) коэффициенты a_i, b_i выражаются через параметры системы и могут быть зависимыми также от неизвестных (искомых) параметров. При этом в общем случае эта зависимость может быть нелинейной.

Ошибка воспроизведения в системе фактическим процессом желаемого или оптимального процесса может быть выражена так:

$$\varepsilon(\varphi, p) = H(\varphi, p) - H_0(p) \doteq \varepsilon(t), \quad (6.8)$$

где $H_0(p)$ — изображение оптимальной переходной функции.

С учетом (6.8) целевая функция может быть представлена в виде

$$J = \int_0^{\omega_m} \gamma(\omega) [H(\varphi, \omega) - H_0(\omega)]^2 d\omega, \quad (6.9)$$

где ω_m — граница частот, в пределах которой требуется осуществить приближение характеристик реального и оптимального процессов по условиям заданной точности; $\gamma(\omega)$ — весовая функция.

При решении задачи синтеза форма желаемого процесса свободного движения может быть любой, например монотонный процесс, процесс, описываемый алгебраической суммой двух экспонент:

$$h_0(t) = H_1 e^{-\alpha_1 t} + H_2 e^{-\alpha_2 t}, \quad (6.10)$$

колебательный затухающий процесс

$$h_0(t) = G e^{-\rho t} \sin(\omega t + \psi), \quad (6.11)$$

где $H_1, H_2, \alpha_1, \alpha_2, \omega, \rho, \psi$ — параметры, определяемые заданными формами переходных и установившихся процессов в системе.

Для задачи оптимизации в качестве оптимального процесса принимается скачкообразный входной сигнал, вследствие чего при подаче на вход постоянного управляющего сигнала оптимальная переходная функция на выходе системы

$$h_0(t) = A_0 1(t), \quad (6.12)$$

где в астатической системе

$$A_0 = h(t) |_{t \rightarrow \infty}; \quad (6.13)$$

в статической системе

$$A_0 = h(t) |_{t \rightarrow \infty} - \Delta; \quad (6.14)$$

Δ — статическая ошибка (допустимая); $1(t)$ — постоянное управляющее воздействие:

$$1(t) = \begin{cases} 0 & \text{при } t < 0, \\ 1 & \text{при } t > 0. \end{cases}$$

Переходные составляющие оптимального переходного процесса равны нулю; имеют место лишь составляющие установившегося режима, которые представляют собой постоянные величины, определяемые выражениями (6.12) — (6.14).

В настоящем параграфе принято преобразование оригиналов функций по Карсону, следовательно, изображение постоянной ве-

личины является также постоянной величиной. На рис. 6.7, б сплошными линиями показаны оптимальные процессы, штриховыми — реальные.

При решении задач синтеза и оптимизации в указанной выше постановке минимизируется функционал (6.9), но при синтезе $H_0(p)$ является изображением конкретного желаемого процесса [например, (6.10) или (6.11)], а при оптимизации в режиме единичного управляющего воздействия $H_0=A_0$ [уравнения (6.12), (6.14)].

Для удобства вычисления целевой функции численными методами заменим в (6.9) интеграл конечной суммой и с учетом (6.7) получим

$$J = \min \sum_{v=0}^{k-1} \gamma(\omega_v) \left[\frac{\sum_{i=0}^m b_i(\varphi) \omega_v^i}{\sum_{i=0}^n a_i(\varphi) \omega_v^i} - H_0(\omega_v) \right]^2, \quad (6.15)$$

где ω_v — узлы аппроксимации характеристики $H_0(\omega)$ характеристикой $H(\varphi, \omega)$; k — число узлов аппроксимации.

Приближенной минимизации целевой функции (6.15) соответствует приближенная минимизация интегральной квадратичной ошибки между соответствующими оригиналами. Для решения задачи используется метод нелинейного программирования.

В качестве ограничений на значения варьируемых параметров принимаются ограничения, обусловленные возможностью физической реализации параметров:

$$\varphi_j^{\min} \leq \varphi_j \leq \varphi_j^{\max}, \quad j = \overline{1, S}. \quad (6.16)$$

Кроме того, вводятся обязательные ограничения по условиям устойчивости системы: $\Delta_i > 0$, где Δ_i — определители Гурвица ($i = \overline{1, n-1}$), n — порядок полинома знаменателя передаточной функции (6.7).

Также вводятся ограничения по значениям колебательности, перегулирования и условиям динамической точности системы автоматического управления.

На рис. 6.8 приведена укрупненная схема алгоритма синтеза и оптимизации линейной системы автоматического управления по интегральной оценке ошибки.

Исходными данными для реализации алгоритма на ЭВМ являются: параметры желаемого (6.10), (6.11) или оптимального (6.12) — (6.14) процесса; изображение переходной функции системы (6.7) с варьируемыми параметрами φ_j ; указанные выше ограничения на определяемые параметры и форму переходного процесса (колебательность, перегулирование, динамическая точность).

В результате решения задачи синтеза определяются оптимальные значения варьируемых параметров φ_j , которые обеспечивают минимум интегральной оценки ошибки воспроизведения требуемого процесса. В результате решения задачи оптимизации определяются оптимальные значения тех же параметров, обеспечивающих

минимум интегральной оценки переходной ошибки. Оба решения должны удовлетворять требованиям устойчивости системы, заданной формы переходного процесса, обеспечения допустимой статической ошибки.

Задача оптимизации нелинейных систем (рис. 6.9) также сводится к определению параметров регулятора из условия минимума интегральной оценки переходной ошибки при обеспечении заданных показателей качества и статической ошибки регулирования. Задача решается методом минимизации переходной ошибки по интегральному квадратичному критерию для обобщенно-линеаризованной системы. Как вычислительная эта задача сводится к многократному решению задачи нелинейно-го программирования.

На рис. 6.9 искомые параметры регулятора φ_j , $j = \overline{1, S}$, входят в линейную часть системы, но они могут являться характеристикой отдельного нелинейного звена. Через $\psi_i(x_i)$, $i = \overline{1, k}$, обозначены нелинейные элементы системы (k — число нелинейных элементов). В качестве рассматриваемого режима работы системы примем режим скачкообразного внешнего воздействия $f(t) = F \cdot 1(t)$, где F — амплитуда. При свободном (невывужденном) движении системы $f(t) = 0$.

Оптимизация рассматриваемой системы по интегральной оценке переходной ошибки выполняется серией итерационных процедур. На каждом шаге итерации производится линеаризация нелинейностей и осуществляется приближенная минимизация интегральной оценки переходной ошибки линеаризованной системы между фактическим и оптимальным процессами, протекающими при синтезированных параметрах φ_j . Линеаризованная задача также решается из условий обеспечения устойчивости и заданных показателей качества управления в системе.

Свободная составляющая выходной величины, преобразованной по Карсону, в линеаризованной системе при постоянном управляющем сигнале в общем случае может быть выражена так:

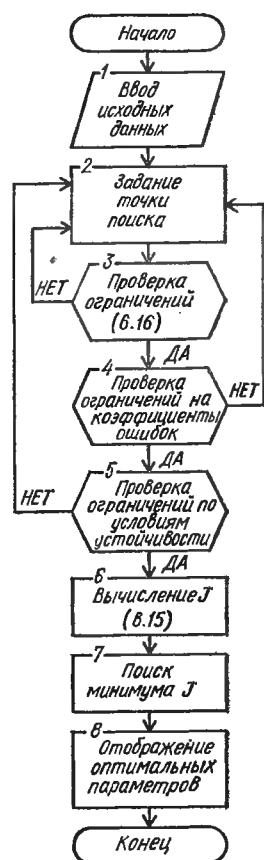


Рис. 6.8. Алгоритм синтеза и оптимизация линейной системы автоматического управления

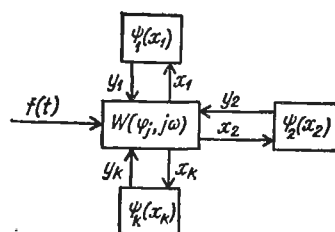


Рис. 6.9. Структура нелинейной системы

$$H(p) = \frac{\sum_{i=0}^m b_i(q, q', \varphi) p^i}{\sum_{i=0}^n a_i(q, q', \varphi) p^i} \doteq h(t), \quad m \leq n, \quad (6.17)$$

где $q = |q_1, q_2, \dots, q_k|$, $q' = |q'_1, q'_2, \dots, q'_k|$ — векторы коэффициентов обобщенной линеаризации нелинейностей.

Преобразованная по Карсону ошибка воспроизведения в системе оптимального процесса фактически имеет такой же вид, как и в линейной системе, и определяется выражением (6.8).

В качестве оптимального процесса $H_0(p)$ принимается переходный процесс (6.12), т. е. скачкообразный выходной сигнал в режиме включения при постоянном управляющем воздействии.

В рассматриваемой постановке задачи в качестве целевой функции принимается интегральная квадратичная оценка погрешности воспроизведения оптимального процесса фактически, т. е.

$$J = \min_{\omega} \sum_{v=0}^{k-1} \gamma(\omega_v) [H(\omega_v) - H_0(\omega_v)]^2, \quad (6.18)$$

где $H(\omega_v)$ — характеристика мнимых частот линеаризованной системы, которая определяется по выражению (6.17) заменой p на $j\omega$ при вещественных положительных значениях ω .

Выполнению условий (6.18) соответствует приближенная минимизация интегральной квадратичной оценки ошибки линеаризованной системы в области оригиналов.

Проанализируем режим свободного движения в рассматриваемой системе, имея в виду, что переходный процесс под воздействием постоянного управляющего сигнала описывается, как и в предыдущем случае.

На первом итерационном шаге нелинейные функции $\Psi_r(x_r)$, $r = \overline{1, k}$, линеаризируются выражениями $\Psi_r(x_r) = (N_{\Psi_r}/2)x_r$. Коэффициент N_{Ψ_r} определяет сектор, внутри которого располагается нелинейная характеристика r -го нелинейного элемента.

Интегральная оценка переходной ошибки минимизируется с помощью варьируемых параметров φ_j . На каждом итерационном шаге задача решается методом нелинейного программирования так же, как для случая линейной системы, рассмотренного выше.

Варьируемые параметры φ_j в первом приближении определяются на первом шаге итерации путем минимизации целевой функции (6.18) при ограничениях, указанных выше. На втором и последующих шагах осуществляется обобщенная линеаризация нелинейных функций. Коэффициенты обобщенной линеаризации q_r , q'_r на втором шаге итерационной процедуры вычисляются на основании связи между выходной величиной системы и входной величиной r -й нелинейности. Эта связь в линеаризованной системе определяется по характеристике первого приближения выходной величины $H_1(p)$, вычисляемой по выражению (6.17) после того, как найдены варьируемые параметры первого приближения φ_j . В результате

получаем некоторый оригинал $h_1(t)$, соответствующий изображению $H_1(p)$.

Переходная характеристика $h_1(t)$ может быть приближенно аппроксимируема переходной характеристикой звена второго порядка:

$$h'_1(t) = ce^{-\rho t} \cos \omega t. \quad (6.19)$$

Преобразование $h'_1(t)$ по Карсону дает

$$H'_1(p) = c \frac{p + \beta_1}{p^2 + 2\beta_1 p + \beta_2}. \quad (6.20)$$

Характеристика $h_1(t)$ аппроксимируется варьированием параметров β_1 и β_2 изображения $H'_1(p)$. Параметры β_1 и β_2 можно определить методом нелинейного программирования, минимизируя целевую функцию

$$J = \min_{\substack{\omega \\ v=0}}^{h-1} \gamma(\omega_v) [H_1(\omega_v) - H'_1(\omega_v, \beta_1, \beta_2)]^2 \quad (6.21)$$

при ограничениях $\beta_1 > 0$, $\beta_2 > 0$ из условий устойчивости.

Расчет переменных $x_r(t)$, $r = \overline{1, k}$, на входах нелинейных элементов по переходной характеристике $h'_1(t)$ может быть выполнен с использованием передаточной функции $S(p)$ линейного звена, расположенного между входом рассматриваемого нелинейного звена и выходом системы. Для этой цели используется выражение

$$\bar{x}_r = S_{h_1 x_r}(p) \bar{h}'_1, \quad (6.22)$$

где $\bar{x}_r$ и $\bar{h}'_1$ — комплексные выражения переменных x_r и h'_1 , причем $\text{mod } \bar{x}_r = A_r$; $\arg \bar{x}_r = \Psi_k$; A_2 — амплитуда; Ψ_k — сдвиг по фазе.

На втором шаге итерации вновь определяются варьируемые параметры φ_j второго приближения для обобщенно-линеаризованной системы, минимизируется целевая функция (6.18) при указанных ограничениях. После определения параметров φ_j второго приближения строится характеристика $H_2(p)$ второго приближения, находятся переменным $h'_2(t)$ и $x_r(t)$ так же, как это выполнялось на первом шаге итерационной процедуры. Итерацион-

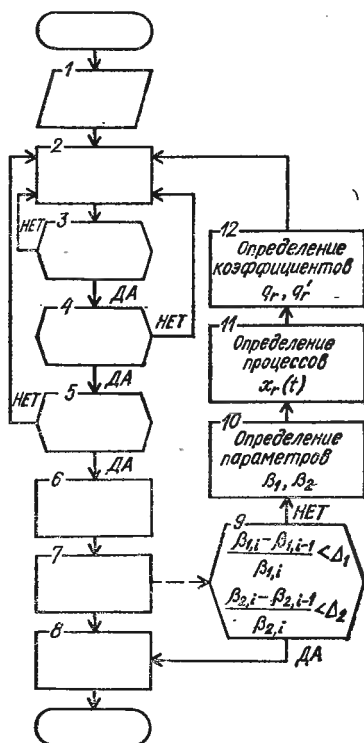


Рис. 6.10. Алгоритм оптимизации нелинейной системы автоматического управления (содержание блоков № 1—8 то же, что и на рис. 6.8)

ный процесс заканчивается после того, как параметры изображения $h'(t)$ аппроксимирующего процесса на выходе системы, т. е. коэффициенты β_1 и β_2 , совпадут с их значениями на предыдущей итерации в пределах заданной точности.

При рассмотрении движения системы под воздействием постоянного управляющего сигнала порядок и характер расчетов аналогичен описанному, но в этом случае вместо (6.19) и (6.20) используется аппроксимирующая функция вида $h'_1(t) = Q_0 - Q_1 e^{-pt} \cos \omega t$.

Схема алгоритма оптимизации нелинейной системы приведена на рис. 6.10, где назначение блоков 1—8 то же, что и на рис. 6.8.

6.4. УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ НА БАЗЕ ЭВМ

Использование цифровых электронно-вычислительных машин в контуре управления динамическими автоматическими системами связано с решением ряда проблем, вытекающих из особенностей ЭВМ как дискретной системы. В связи с этим в САУ с ЭВМ приходится решать вопросы связи ЭВМ с объектом управления и вопросы работы ЭВМ в реальном масштабе времени, т. е. в ритме работы объекта управления.

Связь ЭВМ с объектами управления особенно усложняется при использовании цифровых машин для управления непрерывными автоматическими системами.

Принципиальная схема такой системы представлена на рис. 6.11. В состав схемы входит объект управления (ОУ) с системой регулирования выходных переменных $X\{x_1, x_2, \dots, x_k\}$ с помощью входных переменных $Y\{y_1, y_2, \dots, y_n\}$. Объект находится под влиянием возмущений $F\{f_1, f_2, \dots, f_l\}$. Аналоговые переменные X на выходе объекта преобразуются в дискретную информацию $Z\{z_1, z_2, \dots, z_m\}$ с помощью аналогово-цифрового преобразователя АЦП. С выхода АЦП сигналы обратной связи по выходным переменным объекта поступают в ЭВМ, где обрабатываются с использованием программ, хранящихся на машинных носителях — магнитных дисках (МД), магнитных лентах (МЛ) и др. Выходные переменные ЭВМ $U\{u_1, u_2, \dots, u_r\}$, являющиеся дискретными величинами, преобразуются с помощью цифроаналогового преобразователя (ЦАП) в аналоговые управляющие воздействия на систему управления объектом.

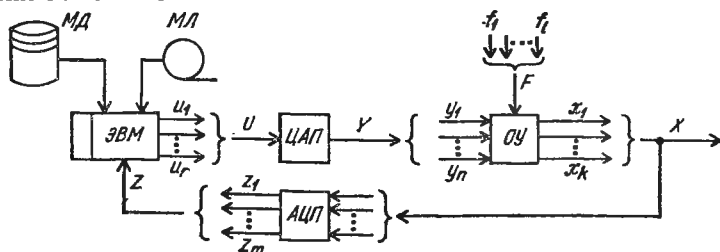


Рис. 6.11. Принципиальная схема автоматического управления с ЭВМ

Как видно из рис. 6.11, проблема связи непрерывной системы управления с ЭВМ решается с помощью аналогово-цифровых и цифроаналоговых преобразователей.

Аналогово-цифровой преобразователь (преобразователь аналог — код) является устройством, осуществляющим автоматическое преобразование (измерение и кодирование) непрерывно изменяющихся во времени аналоговых величин в эквивалентные значения числовых кодов. Количественная связь между аналоговой величиной $x(t_i)$ и соответствующей цифровой величиной z_{t_i} для любого момента времени определяется так:

$$z_{t_i} = \frac{x(t_i)}{\Delta x} \pm [\delta z_{t_i}],$$

где Δx — шаг квантования — аналоговый эквивалент единицы младшего разряда кода; δz_{t_i} — погрешность преобразования на данном шаге.

В качестве входных аналоговых величин $x(t)$ обычно используются временные интервалы, углы поворота, электрические напряжения или токи, частота колебаний, фазовые сдвиги. Выходные коды z_i представляются в двоичной, двоично-десятичной или десятичной системах счисления. Важной характеристикой АЦП является количество каналов, определяющее максимальное число датчиков аналоговых величин, которые могут быть одновременно подключены к преобразователю.

Цифроаналоговый преобразователь (преобразователь код — аналог) является устройством, осуществляющим автоматическое декодирование входных величин, представляемых числовыми кодами, в эквивалентные им значения какой-нибудь физической величины. Количественная связь между входной числовой величиной u_i и ее аналоговым эквивалентом $y(t_i)$ выражается соотношением

$$y(t_i) = u_i \Delta y \pm [\delta y_i],$$

где Δy — аналоговый эквивалент единицы младшего разряда кода; δy_i — погрешность преобразования. Коды u_i представляются в двоичной, двоично-десятичной или десятичной системе счисления. Выходные физические величины $y(t_i)$ — это временные интервалы, угловые перемещения, электрические напряжения или токи, частота колебаний и фазовые сдвиги.

Основными характеристиками АЦП и ЦАП являются точность и быстродействие. Они в значительной степени определяются используемыми в преобразователях элементами — эталонными источниками напряжения и тока, делителями напряжения и тока, ключами и переключателями аналоговых сигналов, устройствами сравнения, усилителями, генераторами линейных или ступенчатых сигналов, устройствами фиксации и запоминания аналоговых сигналов.

ЭВМ в системе автоматического управления осуществляет обработку информации о состоянии объекта, обеспечивает программное и оптимальное управление объектом. На рис. 6.12 представле-

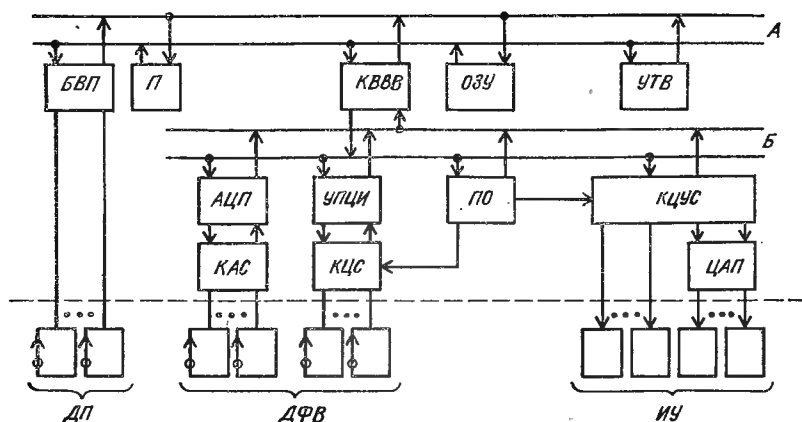


Рис. 6.12. Автоматическое управление технологическим процессом на базе ЭВМ

на принципиальная схема автоматического управления технологическим процессом на базе ЭВМ.

Вся совокупность устройств, подсоединенная к интерфейсу Б, образует систему связи ЭВМ с объектом. Состояние объекта характеризует информация, поступающая от датчиков физических величин (ДФВ). Эта информация после преобразования в соответствующих звеньях системы связи с объектом, показанных на рисунке 6.12, поступает в ЭВМ и составляет поток измерительной информации. От ЭВМ на входы исполнительных устройств (ИУ), приводящих объект в требуемое состояние, поступает поток управляющей информации в виде цифровых, двухпозиционных или аналоговых сигналов. Управляющая информация с каналов ввода-вывода (КВВВ) поступает в коммутатор цифровых управляющих сигналов (КЦУС), с которого передается либо непосредственно на исполнительные устройства дискретного типа, либо в цифроаналоговый преобразователь (ЦАП) для преобразования и последующей передачи на входы аналоговых исполнительных устройств.

К интерфейсу А вместе с каналами ввода-вывода подключен блок внешних прерываний (БВП) процессора (П) и устройство текущего времени (УТВ) — электронные часы. БВП по сигналам от датчиков прерывания (ДП), связанных с объектом управления, и сигналам от УТВ формирует различные циклы обработки информации и управления объектом.

Коммутатор аналоговых сигналов (КАС) и коммутатор цифровых сигналов (КЦС), связанный с устройством приема цифровой информации (УПЦИ), составляют входной коммутатор системы управления, который осуществляет раздельное во времени подключение датчиков. Коммутатор цифровых управляющих сигналов (КЦУС) является выходным коммутатором, осуществляющим раздельное во времени подключение входов исполнительных устройств.

Исполнительные устройства используют либо аналоговые сиг-

налы, либо дискретные, причем длительность дискретных сигналов обычно значительно превышает длительность сигналов управляющей информации. В связи с этим система связи ЭВМ с объектом управления должна содержать технические средства, запоминающие управляющую дискретную информацию до замены ее новой информацией или формирующие управляющие воздействия, воспринимаемые входами исполнительных устройств.

Связь ЭВМ с объектом в системе автоматического управления может быть синхронной, асинхронной и комбинированной.

При *синхронной связи* процесс управления с помощью тактовых сигналов устройства текущего времени (УТВ) разбивается на циклы одинаковой продолжительности. Каждый цикл начинается с появлением тактового сигнала на входе блока прерывания. В начале цикла осуществляется последовательный опрос датчиков, контролирующих состояние объекта управления. Сигналы датчиков преобразуются в форму, необходимую для ввода этих сигналов в ЭВМ. Поступившая в ЭВМ информация обрабатывается и вырабатываются требуемые управляющие воздействия на объект, которые после соответствующего преобразования в системе связи ЭВМ с объектом передаются на исполнительные устройства. Затем ЭВМ либо останавливается, либо выполняет другие программы, не связанные с системой автоматического управления. Выполнение этих программ прерывается следующим тактовым сигналом УТВ. Управляющие воздействия, сформированные в начале цикла, остаются неизменными в течение всего цикла.

При *асинхронной связи* с объектом ЭВМ реагирует на сигналы прерывания, поступающие от датчиков прерывания, непосредственно связанных с объектом. Каждому сигналу прерывания соответствует переход ЭВМ к выполнению соответствующей программы, определяемой характером прерывания. Сигналы прерывания обрабатываются ЭВМ с учетом уровня приоритета.

При *комбинированной связи* ЭВМ с объектом управление осуществляется как по тактовым сигналам УТВ, так и по сигналам прерывания, например по сигналам аварийного режима объекта управления.

В ряде случаев целесообразно использовать прямое цифровое управление объектом на базе ЭВМ. В этих случаях ЭВМ выполняет функции регулятора контура управления. Датчик заданных значений величин, сложение задающих сигналов, сигналов обратных связей и регулятора реализуются в виде программных алгоритмов, а устройство сбора и регистрации переменных состояния и выдачи управляющих воздействий — в виде программируемого функционального устройства. На рис. 6.13, а представлена схема взаимодействия аппаратных и программных средств в системе прямого цифрового управления объектом.

На выходе объекта управления (ОУ), а следовательно, на входе вычислительной системы элементами системы сбора и регистрации информации осуществляется контроль переменных, характеризующих управляемый процесс. Результаты анализа состояния процесса (объекта) X_y с выхода устройства ввода данных управ-

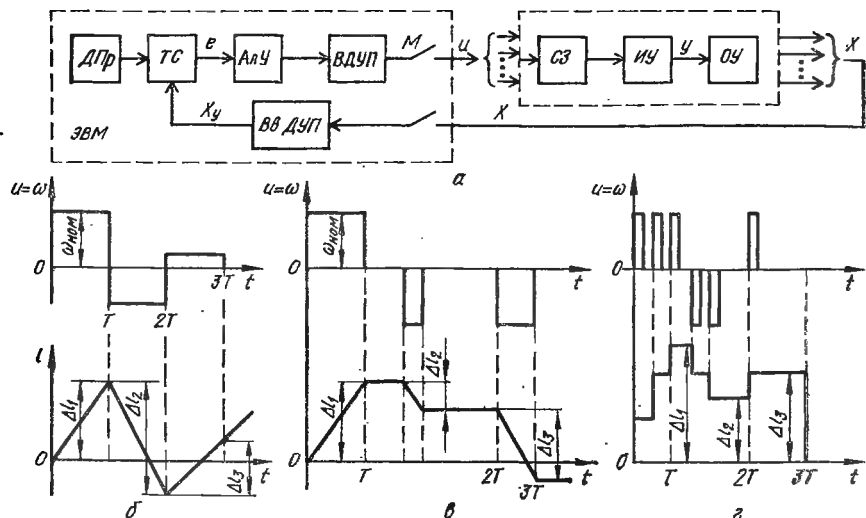


Рис. 6.13. Взаимодействие аппаратных и программных средств в системе прямого цифрового управления:

а — функциональная схема; б — управление с использованием амплитудно-импульсной модуляции; в — управление с использованием временно-импульсной модуляции; г — управление с использованием числа импульсов

ляемого процесса (ВвДуп) поступают в точку смещения (ТС), где с учетом программ, генерируемых датчиком (ДПР), преобразуются в сигнал регулирования $e(t)$, определяющий соответствующий алгоритм управления (АлУ) и выработку сигнала управления u на выходе устройства вывода данных управляемого процесса (ВДУП). Цифровые управляющие сигналы u передаются в соответствующие каналы контуров управления через мультиплексор (М) и заносятся в регистр ЭВМ.

ЭВМ выдает на каждый контур управления управляющее воздействие $\Delta u_k = u_k - u_{k-1}$. При этом цифровые сигналы преобразуются в аналоговые для исполнительного органа и фиксируются в схеме задержки (СЗ). Функция задержки осуществляется для выдачи процессорных данных.

Между управляющим воздействием на объект управления $y(t)$ и переменной $e(t)$ имеет место следующая зависимость:

$$y(t) = k \left[e(t) + \frac{1}{T_{\text{и}}} \int e(t) dt + T_{\text{з}} \frac{de(t)}{dt} \right] \quad (6.23)$$

где k — коэффициент усиления; $T_{\text{и}}$ — время изодрома; $T_{\text{з}}$ — время задержки.

Величина $y(t)$ определяет либо положение l выходного органа объекта управления, либо частоту вращения ω выходного вала. В соответствии с этим реализуется либо алгоритм определения положения

$$l = k \left(e + \frac{1}{T_{\text{и}}} \int e dt + T_{\text{з}} \frac{de(t)}{dt} \right),$$

либо алгоритм определения частоты вращения

$$\omega = \frac{d}{dt} \left[k \left(e + \frac{1}{T_n} \int e dt + T_s \frac{de(t)}{dt} \right) \right]$$

На рис. 6.13, б показано изменение положения l в последовательные моменты времени опроса $T, 2T, \dots$ после того, как сформирован сигнал управления $u = \omega_{\text{ном}}$, где T — промежуток времени между двумя соседними опросами.

На рис. 6.13, в для управления требуемым положением регулируемого органа использована импульсно-временная модуляция, а в случае, представленном на рис. 6.13, г, регулирование осуществляется генерированием определенного числа импульсов для пошаговой установки двигателя.

Алгоритмы прямого цифрового управления могут быть построены подобно алгоритмам аналогового регулирования, разработанным на основе заданной структуры системы и требуемых показателей качества регулирования. При достаточно малом времени опроса ЭВМ, используя схему задержки (СЗ), может с достаточной точностью обеспечить моделирование соответствующего аналогового регулятора.

Рассмотрим формирование цифрового алгоритма управления на основе аналогового алгоритма (6.23). Непрерывные функции $e(t)$ и $y(t)$ при цифровом управлении определяются своими значениями $e_n, e_{n-1}, \dots, y_n, y_{n-1}, \dots$ только в моменты времени опроса $kT, k=1, n$. На основании (6.23) в каждый момент времени управляющий сигнал

$$y_n = k \left(e_n + \frac{1}{T_n} \sum_{k=0}^n e_k T + T_s \frac{e_n - e_{n-1}}{T} \right) \quad (6.24)$$

Значение y_n рассчитывается в каждый момент времени на основании значений x_n или e_n и предшествующих им значений x_{n-1} или e_{n-1} .

Алгоритмы прямого цифрового управления представляют собой дискретные аналоги дифференциальных уравнений аналогового регулирования. Интервал опроса состояния системы выбирается таким, чтобы цифровой регулятор был квазианалоговым. Квазианалоговый алгоритм часто имеет форму

$$y_n = \sum_{k=1}^m b_k y_{n-k} + \sum_{k=0}^m c_k e_{n-k},$$

где b_k и c_k — постоянные.

Прямое цифровое управление позволяет реализовать также алгоритмы, которые технически трудно преобразовать в аналоговую форму. Кроме того, цифровые регуляторы значительно расширяют типы возможных структур систем управления динамическими объектами.

На рис. 6.14 приведена система прямого цифрового управления электроприводом.

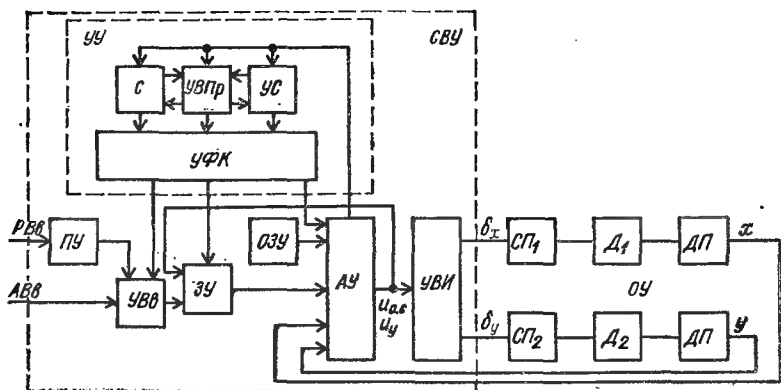


Рис. 6.14. Система автоматического цифрового управления электроприводом

САУ используется для управления двухкоординатной системой воспроизведения движения, например, в периферийных устройствах вычислительных машин, в двухкоординатных станках с числовым программным управлением и т. п.

Объект управления (ОУ) представляет собой два малоинерционных электродвигателя — D_1 , D_2 с полым якорем, подключенных на силовые преобразователи $СП_1$ и $СП_2$, которые выполнены как транзисторные импульсные усилители мощности. Двигатели оснащены датчиками положения (ДП), на выходе которых фиксируются значения управляемых переменных — координат x и y . Эти значения поступают на вход арифметического устройства (АУ), где обрабатываются вместе с выходными сигналами запоминающего устройства (ЗУ), оперативного запоминающего устройства (ОЗУ) и устройства формирования команд (УФК). На вход ЗУ поступают сигналы с выходов АУ, УФК и устройства ввода УВВ. На вход УВВ поступают сигналы с УФК и пульта управления (ПУ), а также автоматически вводимый сигнал управления (АВВ). Устройство управления (УУ) получает сигналы САУ и передает их на устройство формирования команд (УФК) через совокупность элементов, включающих в себя синхронизатор (С), устройство выборки программ (УВПр) и устройство сравнения (УС).

Специализированное вычислительное устройство (СВУ) в целом осуществляет следующие функции: вырабатывает и преобразует управляющие воздействия; реализует выбранный закон управления; отрабатывает сигналы обратных связей по управляемым координатам. Эти функции выполняются с помощью устройств, входящих в СВУ. В арифметическом устройстве (АУ) осуществляются операции сложения, вычитания, умножения. Текущие координаты задающих сигналов, обратных связей и результаты промежуточных вычислений хранятся в ОЗУ, содержимое которого в течение отработки одного кадра многократно меняется. В ЗУ хранятся значения параметров объекта управления, информация о содержании текущего и следующего кадров, постоянные для отрабатываемого

кадра, результаты промежуточных вычислений. УВ и ПУ обеспечивают ручной (РВв) и автоматический (АВв) ввод информации.

Устройство управления вырабатывает команды управления вводом, обработкой и выводом информации. Устройство вывода (УВИ) преобразует информацию, поступающую на его вход с АУ, в широко-импульсные управляющие сигналы δ_x и δ_y , используемые для управления силовыми преобразователями СП₁, СП₂. СВУ выполняется на интегральных микросхемах. В этом устройстве жестко заложены программы: 1) подготовка; 2) интерполяция; 3) расчет управляющих воздействий δ_x ; 4) расчет управляющих воздействий δ_y .

Программа ПОДГОТОВКА осуществляет установку исходного состояния, ввод следующего кадра и расчет коэффициентов закона управления координатами x и y . Программа реализуется один раз в начале каждого кадра.

Программа ИНТЕРПОЛЯЦИЯ управляет движением систем в режимах разгона, равномерного движения, торможения, ползучей скорости. Программа реализуется после появления сигнала интерполяции и заканчивается с обработкой электроприводом координат конечной точки.

Программы РАСЧЕТ δ_x и РАСЧЕТ δ_y реализуются после появления сигналов x и y .

Прямое цифровое управление позволяет создавать системы автоматического управления с высокими динамическими показателями. Например, в рассмотренной схеме управления воспроизведением движения диапазон регулирования частоты вращения достигает $1 \cdot 10^5$, время обработки ступенчатых воздействий около 0,01 с, время перехода с максимальной частоты вращения до остановки с точностью до 1 мкм — около 0,1 с. Создание электромеханической системы автоматического управления с такими высокими показателями качества управления на основе традиционной элементной базы невозможно.

Сравнивая схему прямого цифрового управления с электромашинной системой управления частотой вращения, рассмотренной в гл. 1 (см. рис. 1.7), следует подчеркнуть отсутствие датчика частоты вращения — тахогенератора. Быстродействующее вычислительное устройство обеспечивает время вычисления управляющих воздействий в пределах интервала неуправляемости импульсного преобразователя.

Жесткая логика управления в схеме, приведенной на рис. 6.14, снижает возможности САУ, выполняемых по этому принципу. В частности, новые требования и задачи приводят к необходимости существенной перестройки или полной замены СВУ. Алгоритм управления должен разрабатываться на самом детальном уровне, вплоть до шагов, отражающих элементарные операции.

Существенно расширяются возможности управления в системах, функционирующих на базе микроЭВМ, с сохранением или даже со снижением стоимости программно-аппаратных средств. Здесь становится возможным использовать все преимущества микропрограммирования, позволяющего реализовать набор машинных ко-

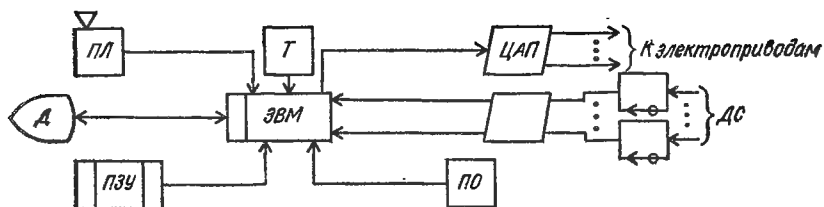


Рис. 6.15. Автоматизация многокоординатного управления динамической системой на базе микроЭВМ

манд стандартных ЭВМ, а также специальные наборы команд, эффективные для определенной области управления. Кроме того, можно реализовать конструкции языка программирования высокого уровня, ядро операционной системы реального времени, диагностические функции для быстрого обнаружения ошибок и сбоев. Возможность распараллеливания на микропрограммном уровне отдельных элементарных операций позволяет значительно повысить быстродействие исполнения алгоритмов.

В качестве примера на рис. 6.15 приведен аппаратный состав микропроцессорного управления многокоординатной динамической системой. В состав системы управления входит микроЭВМ с постоянным запоминающим устройством (ПЗУ), устройством отображения — дисплеем (Д), системой ввода и вывода информации датчиков сигналов (ДС). Управляющая программа может вводиться с перфоленты (ПЛ), с клавиатуры дисплея или с пульта оператора (ПО). С помощью дисплея можно осуществлять коррекцию программы в памяти, задание системных параметров, расширенный диалог с системой.

Возможности системы в значительной степени определяются возможностями входного языка. Управляющая программа обеспечивает обработку контура, состоящего из отрезков прямых и дуг окружностей. Возможно управление вручную в режиме диалога с оператором.

Отметим особенности постоянного запоминающего устройства, использованного в схеме (рис. 6.15). ПЗУ — это память, позволяющая только прочитать то, что в ней записано, и не допускающая возможности хотя бы частичного изменения информации, занесенной в нее. Основное достоинство ПЗУ — неразрушаемость: хранящая в ПЗУ информация не разрушается при выключении питания. В ПЗУ хранятся неизменные программы: микропроцессор только считывает команды из ПЗУ и организует их выполнение, но не записывает новой информации в эту область памяти. Программа управления автоматической системой фиксируется в ПЗУ при условиях, не допускающих случайного или умышленного изменения информации, содержащейся в ПЗУ.

Примером эффективного входного языка систем рассматриваемого типа может служить язык КОНТРАН, разработанный в лаборатории электромеханических систем воспроизведения движения Новосибирского электротехнического института (НЭТИ). Язык

КОНТРАН предназначен для программирования обработки деталей на фрезерных пятикоординатных станках. Средства языка позволяют описать контуры, включающие отрезки и дуги окружностей, расположенных в различных плоскостях, задать контурные скорости для прохождения различных участков контура, обеспечить построение эквидистанты и выполнение ряда технологических функций.

Глава 7

МАТЕМАТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ СЛОЖНЫМИ СИСТЕМАМИ

7.1. ОБЩАЯ ХАРАКТЕРИСТИКА. ОСНОВЫ МЕТОДА ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Задачи управления сложными системами являются типичными оптимизационными задачами, которые описываются функциональными зависимостями (моделями), сложными с математической точки зрения, но простыми по структуре:

$$f_0 = f_0(\bar{X}, \bar{Y}); \quad (7.1)$$

$$f_i = f_i(\bar{X}, \bar{Y}) (\leq, =, \geq) 0, \quad i = \overline{1, m}. \quad (7.2)$$

Здесь f_0 — полезность или значение критерия, характеризующее качество функционирования системы (всегда скаляр); $\bar{X}$ — вектор переменных, которыми нужно управлять; $\bar{Y}$ — вектор переменных (и постоянных), не поддающихся управлению, но влияющих на f_i . Соотношения (7.2) выражают тот факт, что управляемые переменные могут изменяться в определенных пределах. Уравнение (7.1), выражающее целевую функцию, совместно с ограничениями (7.2) образуют модель системы (задачи), которую необходимо решить.

Модель используют для отыскания точных или приближенных оптимальных значений управляемых переменных, т. е. таких значений, которые обеспечивают наилучший показатель качества функционирования системы при заданных значениях неуправляемых переменных. Как именно получается решение задачи, зависит от характера модели.

Модели могут быть линейными, если соотношения (7.1), (7.2) — линейные функции, и нелинейными, если эти функции нелинейные. Для решения оптимизационных задач (7.1), (7.2) существуют различные методы. Независимо от метода, всегда находится оптимальное или близкое к нему решение. Оптимальным является решение, которое минимизирует или максимизирует (в зависимости от сущности задачи) критерий качества (целевую функцию) на модели при заданных условиях и ограничениях, представленных в ней. При этом следует учитывать степень адекватности математической модели описываемой задаче. Вследствие того что модель никогда

точно не описывает задачу, полученное таким путем оптимальное решение никогда не является единственным наилучшим решением реальной задачи, а является аппроксимацией ее оптимального решения.

Если ограничения (7.2) представляют собой строчки равенства и в модель входит небольшое число управляемых переменных, то для отыскания решения часто можно использовать все классические методы математики, например дифференциальное исчисление или уравнения в конечных разностях. Если же эти методы оказываются неспособными привести к результату из-за отсутствия замкнутого аналитического выражения для f_0 вследствие сложности ограничений или большого числа переменных (большой размерности задачи), то прибегают к какому-либо итеративному методу математического программирования, являющемуся составной частью теории принятия решения — исследования операций. Сущность такого метода заключается в том, что вычислительный процесс начинают с некоторого пробного (допустимого) решения, а затем применяют алгоритм, обеспечивающий улучшение этого решения. Исходное пробное решение заменяют улучшенным решением, и процесс продолжают до тех пор, пока не станет ясно, что либо дальнейшее улучшение решения невозможно, либо «стоимость» дальнейших вычислений слишком высока. Ввиду громоздкости и большой размерности такие методы реализуют на ЭВМ.

В качестве метода решения оптимизационных задач используют линейное программирование (ЛП), представляющее собой раздел математического программирования, изучающий задачи нахождения экстремума линейных функций при линейных ограничениях в виде равенств или неравенств.

Линейное программирование — наиболее изученный и широко применяемый метод исследования операций. Впервые идеи линейного программирования были изложены советским математиком Л. В. Канторовичем в 1939 г. в работе «Математические методы организации и планирования производства».

В 1947 г. американский ученый Дж. Данциг опубликовал симплекс-метод решения оптимизационных задач, описываемых линейными моделями, а сам термин «линейное программирование» впервые появился в 1951 г.

Общая задача ЛП описывается математической моделью следующего вида:

максимизировать (минимизировать) линейную функцию

$$f_0(\bar{X}) = \sum_{j=1}^n c_j x_j \quad (7.3)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j = a_{i0}, \quad i = 1, \dots, r; \quad (7.4)$$

$$\sum_{j=1}^n a_{ij} x_j \leq a_{i0}, \quad i = r + 1, \dots, l; \quad (7.5)$$

$$\sum_{j=1}^n a_{ij}x_j \geq a_{i0}, \quad i = l + 1, \dots, m; \quad (7.6)$$

$$x_j \geq 0, \quad j = 1, \dots, n, \quad (7.7)$$

т. е. необходимо найти точку $\bar{X} = (x_1, \dots, x_n)$ n -мерного пространства, которая удовлетворяет системе ограничений (7.4) — (7.7) и максимизирует (минимизирует) целевую функцию (7.3). Иногда условие (7.7) отсутствует.

С организационно-экономической точки зрения модель (7.3) — (7.7) можно охарактеризовать так: есть возможность производить несколько видов продукции ($j = \overline{1, n}$); для производства единицы продукции каждого вида (j) необходимо затратить различные ресурсы $a_{ij}, i = \overline{1, m}$; количество имеющихся ресурсов каждого вида ограничено и равно $a_{i0}, i = \overline{1, m}$; удельная прибыль от произведенной продукции или стоимость реализации ее равна c_j . Необходимо определить объемы (уровни) производимой продукции x_j , при которых оптимизируется (максимизируется или минимизируется) результат производственной деятельности системы в целом без нарушения ограничений, накладываемых на использование ресурсов.

Линейные модели строятся при следующих допущениях:

1) пропорциональность — затраты ресурсов на производство вида продукции и вклад производственного продукта в целевую функцию прямо пропорциональны его уровню;

2) аддитивность — общая сумма ресурсов, расходуемых в системе при производстве всех видов продукции, равна сумме затрат ресурсов на отдельные виды продукции; общая прибыль (стоимость) складывается из частных прибылей (стоимостей), получаемых при реализации (производстве) каждого вида продукции;

3) неотрицательность — объемы произведенной продукции не могут характеризоваться отрицательными значениями.

Два первых допущения обеспечивают строгую линейность целевой функции (7.3) и ограничений (7.4) — (7.6).

Линейное программирование широко применяется при управлении сложными системами (отраслью промышленности, производственными процессами, запасами, транспортом, производственными предприятиями и т. п.).

При описании задач динамического характера коэффициенты модели (7.3) — (7.7) являются функциями некоторых параметров. Часто линейное программирование применяется для решения задач динамического характера, в которых результат предыдущего периода управления определяет исходную информацию последующего. Иногда линейное программирование используется для решения задач, содержащих информацию вероятностного характера. В этих случаях описание соответствующих случайных величин обычно сводят к эквивалентным детерминированным соотношениям.

Линейное программирование используется при решении задач статистики, при доказательстве основных теорем теории игр. Очень часто задачи, которые описываются нелинейными моделями, ап-

проксимируют линейными моделями и получают достаточно хорошие практические результаты.

Особым классом моделей линейного программирования являются транспортные модели, которыми описываются задачи выбора наиболее экономного плана перевозок сырья (продукта, изделий, товаров и др.) от m источников (производителей) к n пунктам назначения (потребителям).

Математическая модель такой задачи имеет вид:
минимизировать

$$f_0(\bar{X}) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (7.8)$$

при ограничениях

$$\sum_{j=1}^n x_{ij} = a_i, \quad i = \overline{1, m}; \quad (7.9)$$

$$\sum_{i=1}^m x_{ij} = b_j, \quad j = \overline{1, n}; \quad (7.10)$$

$$x_{ij} \geq 0, \quad i = \overline{1, m}; \quad j = \overline{1, n}, \quad (7.11)$$

где a_i — возможный объем поставок от i -го источника; b_j — спрос j -го потребителя; c_{ij} — стоимость перевозки единицы сырья i -го источника к j -му потребителю; x_{ij} — объем перевозок от источника i к потребителю j .

Предполагается, что перевозимое сырье (продукт, изделия, товары и др.) одного вида.

В постановке задачи (7.8) — (7.11) требуется выполнение условия $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$, что всегда можно сделать, добавив фиктивный источник или пункт потребления с тем, чтобы компенсировать разность $|\sum_{i=1}^m a_i - \sum_{j=1}^n b_j|$.

Обобщением задачи (7.8) — (7.11) являются транспортная задача с промежуточными пунктами и транспортная задача с ограниченными пропускными способностями. Эти задачи удобно рассматривать на сети.

К задачам транспортного типа относятся:

задача календарного планирования производства, в которой для каждого из плановых периодов определяется оптимальный объем производства, удовлетворяющий плановый спрос;

задача о назначениях, в которой необходимо распределить m исполнителей по n видам работ. Эффективность использования каждого исполнителя на равных работах неодинакова.

Рассмотрим примеры построения моделей задач линейного программирования.

Пример 1. При производстве бетона для строительства используют четыре вида сырья: S_1, S_2, S_3, S_4 ; стоимость единицы каждого вида сырья соответственно c_1, c_2, c_3, c_4 . Бетон должен содержать компоненты трех видов. Количество компонента первого вида должно быть не менее b_1 единиц, второго — b_2 единиц,

третьего — b_3 единиц. Известно, что в единице сырья S_j содержится a_{1j} единиц первого компонента, a_{2j} единиц второго компонента и a_{3j} единиц третьего компонента ($j = 1, 2, 3, 4$).

Требуется так произвести бетон, чтобы его стоимость была минимальной при выполнении заданных условий.

Решение. Пусть x_1, x_2, x_3, x_4 обозначают искомые количества сырья видов S_1, S_2, S_3, S_4 , входящих в бетон.

Общая стоимость бетона

$$L = \sum_{j=1}^4 c_j x_j.$$

Запишем условия содержания в бетоне каждого из компонентов. Если в одной единице сырья вида S_1 содержится первого компонента a_{11} , то в x_1 единицах — $a_{11}x_1$; в x_2 единицах сырья вида S_2 — $a_{12}x_2$ и т. д. Общее количество первого компонента в бетоне не должно быть меньше b_1 ; математически это условие запишется в виде неравенства

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \geq b_1.$$

Аналогично записываются условия для других компонентов:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 &\geq b_1; \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 &\geq b_2; \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &\geq b_3. \end{aligned} \quad (7.12)$$

Эти неравенства являются ограничениями, накладываемыми на искомое решение задачи x_1, x_2, x_3, x_4 .

Таким образом, возникает следующая задача: найти такие неотрицательные значения переменных x_1, x_2, x_3, x_4 , удовлетворяющие линейным ограничениям, при которых линейная функция (целевая функция) этих переменных обращалась бы в минимум:

$$L = \sum_{j=1}^4 c_j x_j \rightarrow \min. \quad (7.13)$$

Соотношения (7.12)–(7.13) описывают математическую модель сформулированной задачи.

Пример 2. Цех столярных изделий домостроительного комбината располагает N_1 станками типа 1, N_2 станками типа 2. Станки могут производить столярные изделия четырех видов: u_1, u_2, u_3, u_4 .

Станки каждого типа могут производить столярные изделия любого вида, но в неодинаковом количестве.

Станок типа i ($i = 1, 2$) производит в месяц a_{i1} единиц изделий u_1 , a_{i2} — изделий u_2 , a_{i3} — изделий u_3 , a_{i4} — изделий u_4 .

Каждая единица изделия u_1 приносит цеху доход c_1 , изделия u_2 — доход c_2 , изделия u_3 — доход c_3 , изделия u_4 — доход c_4 . Цеху предписан план, согласно которому он обязан произвести за месяц: не менее b_1 изделий u_1 , не менее b_2 изделий u_2 , не менее b_3 изделий u_3 , не менее b_4 изделий u_4 .

Требуется так распределить загрузку станков производством изделий различных типов, чтобы план был выполнен и при этом месячная прибыль была максимальной.

Решение. Запишем условия задачи математически (построим математическую модель задачи). Обозначим x_{11} — число станков типа 1, занятых производством изделий u_1 , x_{12} — число станков типа 1, занятых производством изделий u_2 , т. е. x_{ij} — количество станков типа i , занятых производством изделий u_j . Первый индекс соответствует типу станков, второй — типу изделий ($i = 1, 2$; $j = 1, 2, 3, 4$).

Таким образом, необходимо определить восемь переменных:

$$x_{11}, x_{12}, x_{13}, x_{14}; \quad x_{21}, x_{22}, x_{23}, x_{24}$$

так, чтобы месячная прибыль была максимальной.

Запишем формулу для вычисления этой прибыли. Так как каждое изделие u_j , $j = 1, 2, 3, 4$, приносит прибыль c_j , то x_{ij} изделий принесут прибыль $c_j x_{ij}$.

Поскольку станок типа 2 тоже может производить изделия u_j , то изделие u_j принесет цеху прибыль

$$c_j^2(x_{1j} + x_{2j}).$$

Общая прибыль

$$L = c_1(x_{11} + x_{21}) + c_2(x_{12} + x_{22}) + c_3(x_{13} + x_{23}) + c_4(x_{14} + x_{24}). \quad (7.14)$$

Требуется определить такие неотрицательные значения переменных x_{ij} , $i = 1, 2$; $j = 1, 2, 3, 4$, чтобы линейная функция от них (7.14) обращалась в максимум. При этом должны выполняться следующие ограничения:

1) ресурсы по станкам не должны быть превышены, т. е. сумма количеств станков каждого типа, занятых производством изделий всех видов, не должна превышать наличного запаса станков:

$$x_{11} + x_{12} + x_{13} + x_{14} \leq N_1; \quad x_{21} + x_{22} + x_{23} + x_{24} \leq N_2; \quad (7.15)$$

2) задания по ассортименту должны быть выполнены (или перевыполнены):

$$\begin{aligned} a_{11}x_{11} + a_{21}x_{21} &\geq b_1; & a_{13}x_{13} + a_{23}x_{23} &\geq b_3; \\ a_{12}x_{12} + a_{22}x_{22} &\geq b_2; & a_{14}x_{14} + a_{24}x_{24} &\geq b_4. \end{aligned} \quad (7.16)$$

Сформулируем задачу: определить такие неотрицательные значения переменных x_{ij} , $i = 1, 2$; $j = 1, 2, 3, 4$, при которых их линейная функция (7.14) обращается в максимум и выполняются ограничения (7.15), (7.16).

Решения $\bar{X} = (x_1, x_2, \dots, x_n)$, удовлетворяющие ограничениям (7.4)–(7.7), называются *допустимыми решениями модели* и образуют допустимое множество (область) решений. Допустимое решение, максимизирующее (минимизирующее) целевую функцию (7.3), называется *решением задачи линейного программирования* (оптимальным планом).

Задача, описанная моделью (7.3)–(7.7), может не иметь ни одного допустимого решения или может иметь: единственное допустимое оптимальное решение; несколько допустимых оптимальных решений; такое допустимое решение, для которого целевая функция является неограниченной, т. е. неограниченно возрастает для задачи максимизации и неограниченно убывает для задачи минимизации.

Для решения задачи (7.3–7.7), казалось бы, можно применить метод простого перебора: найти произвольное допустимое решение $\bar{X}^{(1)}$, вычислить значение целевой функции $f_0(\bar{X}^{(1)})$, затем найти следующее допустимое решение $\bar{X}^{(2)}$, вычислить для него $f_0(\bar{X}^{(2)})$, сравнить $f_0(\bar{X}^{(1)})$ с $f_0(\bar{X}^{(2)})$, запомнить лучшее значение f_0 и соответствующее допустимое решение; такую процедуру повторять для всех возможных допустимых решений. Значение $\bar{X} = (x_1, \dots, x_n)$, при котором $f_0(\bar{X})$ примет максимальное (минимальное) значение, будет решением задачи линейного программирования. Но в силу сложности поиска и невозможности перебора всех допустимых решений такой путь нереализуем.

Допустимое множество решений задач линейного программирования всегда является многогранным и непустым множеством, если система ограничений (7.4)–(7.7) совместна, т. е. обладает хотя бы одним решением. Если допустимое множество решений не пусто, то оно является выпуклым многогранным множеством. Поэтому задача линейного программирования представляет собой отыскание максимума (минимума) линейной функции на многогранном выпуклом множестве. С геометрической точки зрения решению задач линейного программирования соответствует одна из вершин множества допустимых решений. Поэтому отыскание решения задачи (7.3)–(7.7) сводится к перебору конечного числа вершин выпуклого многогранного множества.

Для наглядности рассмотрим задачу линейного программи-

рования с m ограничениями и двумя управляемыми переменными:

$$\max (c_1x_1 + c_2x_2); \quad (7.17)$$

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 &\leq a_{10} \\ a_{m1}x_1 + a_{m2}x_2 &\leq a_{m0} \end{aligned} \right\} \quad (7.18)$$

$$x_1, x_2 \geq 0. \quad (7.19)$$

Пусть система (7.18) совместна и соответствующая область допустимых решений ограничена, т. е. является многоугольником. Каждое из ограничений (7.18) определяет полуплоскость соответственно с граничной прямой $a_{i1}x_1 + a_{i2}x_2 = a_{i0}$, $i = \overline{1, m}$.

Если функцию цели приравнять некоторой константе ζ , то получим тоже уравнение прямой

$$c_1x_1 + c_2x_2 = \zeta.$$

Изменяя константу ζ , получаем семейство параллельных прямых — линии уровня целевой функции.

Пусть область допустимых решений (выпуклый многоугольник) D и линии уровня целевой функции соответствуют рис. 7.1.

Если прямую $c_1x_1 + c_2x_2 = \zeta$ перемещать в направлении ее нормали $S = (c_1, c_2)$, то значение ζ будет неограниченно возрастать. Линии уровня вначале могут не иметь общих точек с областью D , а затем при некотором ζ прямая $c_1x_1 + c_2x_2 = \zeta$ коснется области D (точка N), потом пересечет ее и, наконец, коснется ее в последний раз в точке M ; в этой точке функция цели достигнет своего максимального значения.

Если область допустимых решений представляет неограниченный многоугольник, то может быть один из двух случаев: прямая $c_1x_1 + c_2x_2 = \zeta$, передвигаясь по нормали, все время пересекает область D или становится опорой к ней. В первом случае целевая функция неограниченно возрастает, во втором — достигает своего максимального значения.

Таким образом, можно утверждать, что целевая функция достигает своего максимального значения в вершине многоугольника (угловой точке). Вершины области допустимых решений называют *экстремальными точками*. Если целевая функция достигает максимального значения более, чем в одной точке, то она достигает этого же значения в любой точке, являющейся их выпуклой комбинацией (прямой, соединяющей эти экстремальные точки).

Следовательно, чтобы решить задачу линейного программирования, необходимо рассматривать только решения, соответствующие угловым точкам области допустимых решений. В связи с этим необходимы правила, следуя

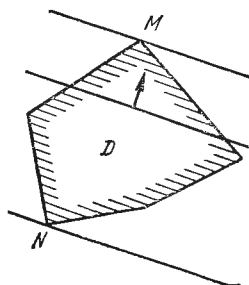


Рис. 7.1. Геометрическая интерпретация задачи линейного программирования

которым можно определять угловые точки и целенаправленно переходить от одной угловой точки допустимой области к другой.

Каноническая форма задач линейного программирования

Существующие методы решения задач линейного программирования применимы к моделям этих задач, представленным в канонической форме, которая характеризуется следующими тремя признаками:

система ограничений представлена в виде системы уравнений;

на все переменные накладываются условия неотрицательности;

производится максимизация (минимизация) целевой функции.

В канонической форме общая задача линейного программирования имеет вид:

максимизировать

$$f_0 = \sum_{j=1}^n c_j x_j \quad (7.20)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j = a_{i0}, \quad i = \overline{1, m}; \quad (7.21)$$

$$x_j \geq 0, \quad j = \overline{1, n}, \quad (7.22)$$

где c_j , a_{ij} , a_{i0} — заданные постоянные величины; x_j — искомые управляемые переменные.

Задача (7.3) — (7.7) переводится в каноническую форму (7.20) — (7.22) путем преобразования всех ограничений, имеющих вид неравенств, в равенства. Это достигается добавлением неотрицательных остаточных переменных в левую часть неравенств (7.5) и вычитанием неотрицательных переменных в левой части неравенств (7.6). Ограничения (7.5) и (7.6) примут вид:

$$a_{r+1,1}x_1 + a_{r+1,2}x_2 + \dots + a_{r+1,n}x_n + 1 \cdot x_{n+1} = a_{r+1,0};$$

$$a_{r+2,1}x_1 + a_{r+2,2}x_2 + \dots + a_{r+2,n}x_n + 0 \cdot x_{n+1} + 1 \cdot x_{n+2} = a_{r+2,0};$$

$$a_{l+1,1}x_1 + a_{l+1,2}x_2 + \dots + a_{l+1,n}x_n + 0 \cdot x_{n+1} + \dots + 0 \cdot x_{n+l-1} + 1 \cdot x_{n+l} = a_{l+1,0};$$

$$a_{l+1,1}x_1 + a_{l+1,2}x_2 + \dots + a_{l+1,n}x_n - 1 \cdot x_{n+l+1} = a_{l+1,0};$$

$$a_{l+2,1}x_1 + a_{l+2,2}x_2 + \dots + a_{l+2,n}x_n - 0 \cdot x_{n+l+1} - 1 \cdot x_{n+l+2} = a_{l+2,0};$$

$$a_{m+1,1}x_1 + a_{m+1,2}x_2 + \dots + a_{m+1,n}x_n - 0 \cdot x_{n+l+1} - \dots$$

$$\dots - 0 \cdot x_{n+l+m-1} - 1 \cdot x_{n+m} = a_{m+1,0}.$$

Таким образом, для приведения задачи (7.3) — (7.7) к каноническому виду необходимо ввести в ограничения (7.5) — (7.6) $m-r$ свободных переменных, из которых $l-r$ — остаточные переменные и $m-l$ — избыточные переменные. Целесообразно также сделать неотрицательными все элементы правой части ограничений a_{i0} . При

наличии переменных, которые не удовлетворяют условиям (7.7), необходимо вместо переменной x_p ввести в модель две неотрицательные переменные x_p^+ и x_p^- :

$$x_p = x_p^+ - x_p^-.$$

В результате преобразования модели (7.3)—(7.7) к канонической форме (7.20)—(7.22) количество управляемых переменных x_j возрастет.

Задача минимизации может быть сведена к эквивалентной задаче максимизации и наоборот, если одновременно с изменением знака оптимизации изменить знаки перед всеми коэффициентами в выражении для целевой функции, т. е. минимизация $\sum_{j=1}^n c_j x_j$ эквивалентна максимизации $\sum_{j=1}^n (-c_j) x_j$. При этом

$$\min \sum_{j=1}^n c_j x_j = \max \sum_{j=1}^n (-c_j) x_j.$$

Приведенную к канонической форме модель задачи линейного программирования часто удобно рассматривать в матричном или векторном представлении.

Матричное представление модели:
минимизировать (максимизировать)

$$x_0 = \bar{C}\bar{X} \quad (7.23)$$

при ограничениях

$$\bar{A}\bar{X} = \bar{A}_0; \quad (7.24)$$

$$\bar{X} \geq 0, \quad (7.25)$$

где $\bar{X} = (x_1, \dots, x_n)$ — вектор, компонентами которого являются управляемые переменные (в число этих компонентов могут входить остаточные и избыточные переменные); $\bar{A}_0 = (a_{10}, \dots, a_{m0})$ — вектор, компонентами которого являются неотрицательные правые части ограничений; $\bar{A} = \|a_{ij}\|_{m \times n}$ — матрица коэффициентов ограничений, включающая коэффициенты при остаточных и избыточных переменных.

Векторное представление модели:
максимизировать (минимизировать)

$$x_0 = \sum_{j=1}^n c_j x_j \quad (7.26)$$

при ограничениях

$$\sum_{j=1}^n \bar{A}_j x_j = \bar{A}_0; \quad (7.27)$$

$$x_j \geq 0, \quad j = \overline{1, n}, \quad (7.28)$$

где $\bar{A} = (\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n)$. Векторы $\bar{A}_j$ и $\bar{A}_0$ часто называют *векторами условий*.

Обычно количество ограничений (7.27) меньше количества управляемых переменных, т. е. $m < n$. Если $m > n$, то система ограничений содержит по крайней мере $m - n$ избыточных уравнений;

если $m=n$ и $\bar{A}$ — матрица полного ранга (все ограничения линейно независимы), то система $\bar{A}\bar{X}=\bar{A}_0$ имеет единственное решение.

В случае $m < n$ система $\bar{A}\bar{X}=\bar{A}_0$ имеет бесконечное множество решений; если эта система несовместна, задача не имеет ни одного допустимого решения.

Допустим система совместна, выберем из матрицы $\bar{A}=(\bar{A}_1, \dots, \bar{A}_n)$ m линейно независимых векторов, обозначив матрицу, содержащую их, через $\bar{B}^{m \times m}$. Эта матрица образует базис системы. Матрица $\bar{A}$ может быть представлена в виде $\bar{A}=(\bar{B}, \bar{N})$, где $\bar{N}$ — матрица, содержащая небазисные векторы.

Ограничения (7.24) переписываются так:

$$\bar{B}\bar{X}_B + \bar{N}\bar{X}_N = \bar{A}_0, \quad (7.29)$$

где $\bar{X}_B$ — вектор базисных переменных, связанных с матрицей $\bar{B}$; $\bar{X}_N$ — вектор небазисных переменных, связанных с матрицей $\bar{N}$. Так как $\bar{B}$ — невырожденная квадратная матрица, то существует обратная матрица $\bar{B}^{-1}$. Выразим базисные переменные через небазисные, умножив обе части соотношения (7.29) на $\bar{B}^{-1}$:

$$\bar{B}^{-1}\bar{B}\bar{X}_B + \bar{B}^{-1}\bar{N}\bar{X}_N = \bar{B}^{-1}\bar{A}_0.$$

Так как $\bar{B}^{-1}\bar{B}=1$, получим

$$\bar{X}_B = \bar{B}^{-1}\bar{A}_0 - \bar{B}^{-1}\bar{N}\bar{X}_N. \quad (7.30)$$

Соотношения (7.30) определяют полное множество решений системы линейных уравнений (7.24).

Небазисные переменные могут принимать произвольные значения. Пусть все небазисные переменные принимают нулевые значения, тогда базисное решение

$$\bar{X}_B = \bar{B}^{-1}\bar{A}_0. \quad (7.31)$$

Если базисное решение $\bar{X}_B \geq 0$, то оно называется *допустимым*.

Если все базисные переменные вектора $\bar{X}_B$ строго положительны, то базисное решение является *допустимым невырожденным*. Если хотя бы одна переменная вектора $\bar{X}_B$ равна нулю, допустимое базисное решение называется *вырожденным*.

Количество экстремальных точек определяется числом сочетаний

$$C_n^m = \frac{n!}{m!(n-m)!}$$

т. е. количеством допустимых базисных решений системы $\bar{A}\bar{X} = \sum_{j=1}^n \bar{A}_j \bar{X}_j = \bar{A}_0$. Невырожденному допустимому базисному решению соответствует невырожденная экстремальная точка, а вырожденному — вырожденная. Невырожденная экстремальная точка определяется только одним базисным решением, вырожденная — двумя и более.

Основная идея симплекс-метода заключается в целенаправленном последовательном нахождении экстремальных точек с помощью допустимых базисных решений системы ограничений. Этот упорядоченный процесс начинается с некоторой начальной экстремальной точки (исходного допустимого базисного решения) и продолжается до тех пор, пока не будет найдена точка, соответствующая оптимальному решению.

В качестве каждой новой экстремальной точки (допустимого базисного решения) выбирается смежная с предыдущей экстремальная точка, в которой, по крайней мере, не ухудшается значение целевой функции. Обратный переход к предшествующей экстремальной точке невозможен. Поэтому в процессе вычислительной процедуры после определения нового допустимого базисного решения предыдущее исключается. Максимальное количество итераций (переходов к смежным экстремальным точкам) не превышает C_n^m [n — количество переменных, а m — количество ограничений задачи (7.20) — (7.22)].

Смежные экстремальные точки отличаются только одной переменной. Поэтому каждую последующую экстремальную точку определяют заменой одной из текущих базисных переменных новой из числа небазисных.

Таким образом, симплекс-метод предполагает следующие процедуры:

1. Определение начального допустимого базисного решения. Это решение получается приравниванием нулю $n - m$ небазисных переменных. Исключение из выражения для целевой функции базисных переменных.

2. Определение из числа небазисных переменных переменной, которая должна быть включена в число базисных переменных. Эта новая базисная переменная должна улучшить значение целевой функции.

3. Определение из числа базисных переменных переменной, которая должна быть исключена из базисного решения, т. е. заменена новой базисной переменной.

Две последние процедуры повторяются до тех пор, пока невозможно улучшить значение целевой функции.

Определение начального допустимого базисного решения осуществляется просто, если в ограничениях (7.24) имеется единичный базис, т. е. они имеют вид:

$$A\bar{X}_N + E\bar{X}_B = \bar{A}_0; \quad (7.32)$$

$$\bar{X}_N \geq 0, \quad \bar{X}_B \geq 0, \quad (7.33)$$

где E — единичная матрица.

Такой вид имеют ограничения (7.24), если при переводе модели задачи в каноническую форму в каждое ограничение была введена остаточная переменная.

Если при переводе модели задачи в каноническую форму в каждое ограничение были введены избыточные переменные или некоторые ограничения исходной модели имели вид равенств, то для приведения ограничений к виду (7.32) в них вводят искусственные переменные таким образом, чтобы организовать единичную матрицу $\bar{E}$. Так как в исходную задачу (7.23)—(7.25) искусственные переменные не входят, то в оптимальном решении задачи они должны равняться нулю. Это условие достигается введением в исходную целевую функцию искусственных переменных с коэффициентами, принимающими большие значения, т. е. введением «больших штрафов». Так преобразованную задачу называют *М-задачей*.

Пример:

$$\begin{aligned} \max \quad & (3x_1 - x_2 + 2x_3 + x_4); & 4x_1 + x_2 - 2x_3 &= 4; \\ & 2x_1 - x_2 + 4x_3 + x_4 = 10; & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \\ & -3x_1 + 2x_2 + x_3 - 2x_4 = 8; \end{aligned}$$

Введя три искусственные переменные y_1, y_2, y_3 , получим *М-задачу*:

$$\begin{aligned} \max \quad & (3x_1 - x_2 + 2x_3 + x_4 - My_1 - My_2 - My_3); \\ & 2x_1 - x_2 + 4x_3 + x_4 + 1 \cdot y_1 = 10; \\ & -3x_1 + 2x_2 + x_3 - 2x_4 + 0 \cdot y_1 + 1 \cdot y_2 = 8; \\ & 4x_1 + x_2 - 2x_3 + 0 \cdot x_4 + 0 \cdot y_1 + 0 \cdot y_2 + 1 \cdot y_3 = 4; \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{aligned}$$

Предполагается, что $M \gg c_j, j = \overline{1, 4}; M > 0$.

Начальное допустимое базисное решение: $y_1 = 10; y_2 = 8; y_3 = 4$.

Таким образом, в случае ограничения вида (7.32) начальное допустимое базисное решение имеет вид

$$\bar{X}_B = \bar{A}_0.$$

Новая базисная переменная определяется по симплекс-критерию: если в выражении для целевой функции (после исключения базисных переменных предыдущей итерации) имеются небазисные переменные, за счет которых можно увеличить значение целевой функции, то в базисное решение вводится та переменная, которая дает наибольшее приращение целевой функции (задача максимизации). Если базисные переменные $\bar{X}_B$ выражены через небазисные $\bar{X}_B = \bar{B}^{-1}\bar{A}_0 - \bar{B}^{-1}\bar{N}\bar{X}_N$, которые исключены из выражения целевой функции, то целевая функция примет вид

$$x_0 = \bar{C}_B \bar{B}^{-1} \bar{A}_0 - (\bar{C}_B \bar{B}^{-1} \bar{N} - \bar{C}_N) \bar{X}_N, \quad (7.34)$$

где $\bar{C}_B$ и $\bar{C}_N$ — векторы, компонентами которых являются соответственно коэффициенты целевой функции при базисных и небазисных переменных.

Так как небазисные переменные $\bar{X}_N = 0$, то каждая итерация имеет допустимое базисное решение

$$\bar{X}_B = \bar{B}^{-1} \bar{A}_0,$$

а целевая функция

$$x_0 = \bar{C}_B \bar{B}^{-1} \bar{A}_0.$$

Очевидно, что значение целевой функции может быть увеличе-

но за счет той переменной, для которой выражение $(\bar{C}_B \bar{B}^{-1} \bar{N} - \bar{C}_N)$ будет отрицательно. Поэтому в новый базис необходимо включать тот вектор $\bar{A}_j$ (ввести новую базисную переменную x_j), для которого разность

$$\bar{C}_B \bar{B}^{-1} \bar{A}_j - c_j \quad (7.35)$$

имеет наибольшее отрицательное значение. Если такого вектора нет, то полученное допустимое базисное решение является оптимальным. Выражение (7.35) называется *симплекс-разностью*.

Чтобы выделить в текущей матрице условий единичную подматрицу, т. е. выразить все базисные переменные через небазисные, пользуются методом полного исключения Жордана — Гаусса. Этот метод основан на эквивалентных преобразованиях системы уравнений с целью разрешения на каждом шаге одного уравнения относительно одной неизвестной (базисной переменной) и исключения ее из всех остальных уравнений, что приводит систему к единичному базису, состоящему из единичных векторов-коэффициентов.

Переменная, выводимая из базиса, определяется из условия, что базис должен все время оставаться допустимым, и при вводе в базис новой переменной одна из его переменных должна принять значение, равное нулю. Поэтому из базиса должен быть выведен вектор $\bar{A}_l$, для которого справедливо соотношение

$$\frac{(\bar{B}^{-1} \bar{A}_0)_l}{(\bar{B}^{-1} \bar{A}_j)_l} = \min_i \left\{ \frac{(\bar{B}^{-1} \bar{A}_0)_i}{(\bar{B}^{-1} \bar{A}_j)_i} \mid (\bar{B}^{-1} \bar{A}_j)_i > 0 \right\}, \quad (7.36)$$

где $(\bar{B}^{-1} \bar{A}_0)_i$ — i -й компонент вектора $\bar{B}^{-1} \bar{A}_0$, $i = \overline{1, m}$.

Если все $(\bar{B}^{-1} \bar{A}_j)_i \leq 0$, то выбор $\bar{A}_j$ для ввода в базис не может привести к допустимому базисному решению.

Реализовать симплексный метод решения задач удобно в виде таблиц, заполняемых на каждой итерации в соответствии с симплексными преобразованиями, которые являются модификацией метода полного исключения Жордана — Гаусса.

Таблица 7.1

C			c_1	c_2	...	c_j	...	c_n
	B_K	A_0	A_1	A_2	...	A_j	...	A_n
c_1	x_1	α_{10}	α_{11}	α_{12}	...	α_{1j}	...	α_{1n}
c_2	x_2	α_{20}	α_{21}	α_{22}	...	α_{2j}	...	α_{2n}
...	...	...	...	...	...	...	...	...
c_i	x_i	α_{i0}	α_{i1}	α_{i2}	...	α_{ij}	...	α_{in}
...	...	...	...	...	...	...	...	...
c_m	x_m	α_{m0}	α_{m1}	α_{m2}	...	α_{mj}	...	α_{mn}
	Δ	α_{00}	Δ_1	Δ_2	...	Δ_j	...	Δ_n

Алгоритм решения задачи методом симплекс-таблиц содержит следующие этапы:

1. Заполнение исходной таблицы параметрами, соответствующиминачальному допустимому базисному решению (табл. 7.1). Такими параметрами являются: базисные переменные $x_i, i \in I$, где I — множество индексов базисных переменных (столбец B_x); базисные переменные a_{i0} (столбец A_0); элементы матрицы $\bar{A} = \|a_{ij}\|_{m,n}$, содержащей допустимый единичный базис (столбцы $A_1, A_2, \dots, A_n$); коэффициенты целевой функции (строка C); коэффициенты целевой функции при базисных переменных (столбец C); оценки векторов условий (симплекс-разности)

$$\Delta_j = a_{0j} = \sum_{i \in I} c_i a_{ij} - c_j, \quad j \notin I, \quad (7.37)$$

где Δ — индексная строка; a_{00} — целевая функция: $a_{00} = \sum_{i \in I} c_i a_{i0}$.

Если все $\Delta_j \geq 0$, то данное допустимое базисное решение является оптимальным; если $\Delta_k < 0$ и все элементы столбца $A_k a_{ik} \leq 0, i = \overline{1, m}$, то задача неразрешима; если в каждом столбце A_k , для которого $\Delta_k < 0$, имеется хотя бы один положительный элемент, то возможен переход к новому допустимому базисному решению, более близкому к оптимальному. В последнем случае осуществляется переход ко второму этапу.

2. Выбор вектора A_p , вводимого в базис, из условия

$$\Delta_p = a_{0p} = \min_{1 \leq l \leq n} \{a_{0l} \mid a_{0l} < 0\}. \quad (7.38)$$

3. Выбор вектора A_q , выводимого из базиса, из условия

$$\frac{a_{q0}}{a_{qp}} = \min \left\{ \frac{a_{r0}}{a_{rp}} \mid a_{rp} > 0 \right\}. \quad (7.39)$$

4. Осуществление одного шага симплекс-преобразования с направляющим элементом a_{qp} , для чего используют формулы:

$$a_{ql}^{(k+1)} = \frac{a_{ql}^{(k)}}{a_{qp}^{(k)}}, \quad l = 0, 1, \dots, n; \quad (7.40)$$

$$a_{rp}^{(k+1)} = 0, \quad r = 1, 2, \dots, m; \quad r \neq q; \quad a_{qp}^{(k+1)} = 1; \quad (7.41)$$

$$a_{rl}^{(k+1)} = a_{rl}^{(k)} - \frac{a_{ql}^{(k)}}{a_{qp}^{(k)}} a_{rp}^{(k)}, \quad l \neq p, \quad r \neq q, \quad (7.42)$$

где k — номер итерации (симплексного преобразования). Элементы индексной строки новой таблицы вычисляют тоже по соотношениям (7.40) — (7.42).

Правильность вычислений элементов индексной строки контролируют по формулам:

$$a_{00}^{(k+1)} = \sum_{i \in I^{(k+1)}} c_i a_{i0}^{(k+1)}; \quad (7.43)$$

$$a_{0l}^{(k+1)} = \sum_{i \in I^{(k+1)}} c_i a_{il}^{(k+1)} - c_l, \quad (7.44)$$

где $I^{(k+1)}$ — множество индексов базисных переменных $k+1$ итерации. В столбце B_x заменяют x_q на x_p , а в столбце C — c_q на c_p .

5. Если все $\Delta_l^{(k+1)} = a_{0l}^{(k+1)} \geq 0$, $l = \overline{1, n}$, то новое базисное решение $x_i = a_{i0}^{(k+1)}$, $i \in I^{(k+1)}$, — оптимальное. В противном случае осуществляют переход ко второму этапу, т. е. выполняют следующую итерацию.

При решении задач минимизации в базис вводят вектор с наибольшей положительной оценкой.

Если в качестве начального базиса выбирают базис из свободных переменных, для которых $c_p = 0$, то оценки для всех небазисных переменных $\Delta_h = a_{0h} = -c_h$, а соответствующее значение целевой функции $a_{00} = \sum_{i \in I} c_i x_i = 0$.

Количество симплекс-итераций N при решении задач линейного программирования зависит в основном от числа ограничений. Обычно $1,5m \leq N \leq 2m$, где m — количество ограничений в исходной симплекс-таблице.

Пример 1. Найти решение следующей задачи:

$$\begin{aligned} \max (2x_1 + x_2 + 7x_3); \quad & x_1 + 2x_2 + 3x_3 \leq 4; \\ -x_1 - 4x_2 + 10x_3 \leq 7; \quad & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{aligned}$$

Преобразуем модель задачи к каноническому виду, введя остаточные переменные x_4 и x_5 соответственно в первое и второе ограничения. Каноническая модель задачи примет вид

$$\begin{aligned} \max (2x_1 + x_2 + 7x_3); \quad & x_1 + 2x_2 + 3x_3 + x_4 = 4; \\ -x_1 - 4x_2 + 10x_3 + x_5 = 7; \quad & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0. \end{aligned}$$

Так как вектор $\bar{A}_0 > 0$, а векторы $\bar{A}_4$ и $\bar{A}_5$ образуют единичный базис, то задачу можно решить методом симплекс-таблиц. Табл. 7.2 — начальная симплекс-таблица.

1 итерация

Определяем направляющий столбец. Им будет столбец A_3 , так как $\Delta_3 < \Delta_1 < \Delta_2$ и $\Delta_3 < 0$: $(-7 < -2 < -1)$. Направляющую строку выбираем из условия

$$\min \left\{ \frac{4}{3}, \frac{7}{10} \right\} = \frac{7}{10},$$

т. е. направляющей является вторая строка таблицы. Поэтому направляющий элемент $x_{23} = 10$, т. е. из числа базисных переменных выводится x_5 , а вводится переменная x_3 .

Таблица 7.2

C			2	1	7	0	0
	B_x	A_0	A_1	A_2	A_3	A_4	A_5
0	x_4	4	1	2	3	1	0
0	x_5	7	-1	-4	10	0	1
	Δ	0	-2	-1	-7	0	0

Таблица 7.3

C			2	1	7	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
0	x_4	$\frac{19}{10}$	$\frac{13}{10}$	$\frac{32}{10}$	0	1	$-\frac{3}{10}$
7	x_3	$\frac{7}{10}$	$-\frac{1}{10}$	$-\frac{4}{10}$	1	0	$\frac{1}{10}$
	Δ	$\frac{49}{10}$	$-\frac{27}{10}$	$-\frac{38}{10}$	0	0	$\frac{7}{10}$

Выполнив первый шаг симплексных преобразований, получим табл. 7.3.

II итерация

Поскольку минимальная оценка $\Delta_2 = \frac{38}{10}$, направляющим столбцом является A_2 . Направляющая строка — первая, так как в столбце A_2 имеется только один положительный элемент, расположенный в первой строке.

Следовательно, направляющий элемент $x_{12} = \frac{32}{10}$, и из числа базисных переменных выводится переменная x_4 , а вводится переменная x_2 .

Проделав симплексные преобразования, получим табл. 7.4.

III итерация

Поскольку в индексной строке Δ последней таблицы только один отрицательный элемент Δ_1 , столбец A_1 является направляющим, и переменная x_1 должна быть введена в число базисных переменных. Направляющей строкой будет первая, так как

$$\min \left\{ \frac{19}{13}, \frac{15}{1} \right\} = \frac{19}{13}$$

Поэтому элемент $x_{11} = \frac{13}{32}$ — направляющий, и из числа базисных переменных необходимо вывести x_2 . Проделав симплексные преобразования, получим табл. 7.5, в которой все элементы индексной строки Δ положительны, что свидетельствует о получении оптимального решения.

Таким образом, оптимальное решение задачи $\bar{X} = \left(\frac{19}{13}, 0, \frac{11}{13} \right)$. Значение целевой функции, соответствующее оптимальному решению $a_{00} = \frac{115}{13}$.

Таблица 7.4

C			2	1	7	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
1	x_2	$\frac{19}{32}$	$\frac{13}{32}$	1	0	$\frac{10}{32}$	$-\frac{3}{32}$
7	x_3	$\frac{15}{16}$	$\frac{1}{16}$	0	1	$\frac{1}{8}$	$\frac{1}{16}$
	Δ	$\frac{229}{32}$	$-\frac{37}{32}$	0	0	$\frac{19}{16}$	$\frac{11}{32}$

Таблица 7.5

C			2	1	7	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
2	x_1	$\frac{19}{13}$	1	$\frac{32}{13}$	0	$\frac{10}{13}$	$-\frac{3}{13}$
7	x_3	$\frac{11}{13}$	0	$-\frac{2}{13}$	1	$\frac{1}{13}$	$\frac{1}{13}$
	Δ	$\frac{115}{13}$	0	$\frac{37}{13}$	0	$\frac{27}{13}$	$\frac{1}{13}$

Не всегда модель задачи имеет единичный базис, т. е. нельзя просто построить начальное допустимое решение. В таких случаях удобно построить искусственный единичный базис, а затем воспользоваться методом симплекс-таблиц.

Пример 2. Найти решение задачи

$$\begin{aligned} \max (3x_1 - x_2 + 2x_3 + x_4); \quad & 2x_1 - x_2 + 4x_3 + x_4 = 10; \\ & -3x_1 + 2x_2 + x_3 - 2x_4 = 8; \quad 4x_1 - x_2 - 2x_3 = 4; \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \end{aligned}$$

Строим M -задачу:

$$\begin{aligned} \max (3x_1 - x_2 + 2x_3 + x_4 - My_1 - My_2 - My_3); \quad & 2x_1 - x_2 + 4x_3 + x_4 + y_1 = 10; \\ & -3x_1 + 2x_2 + x_3 - 2x_4 + y_2 = 8; \quad 4x_1 - x_2 - 2x_3 + y_3 = 4; \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{aligned}$$

Исходная симплекс-таблица — табл. 7.6.

I итерация

Наименьшей является оценка $\Delta_1 = -3M - 3$. Поэтому переменная x_1 должна быть введена в число базисных. Так как $\min \left\{ \frac{10}{2}, \frac{4}{4} \right\} = \frac{4}{4}$, то из числа базисных переменных выводится y_3 . Проведав симплексные преобразования с направляющим элементом x_{31} , получим табл. 7.7.

II итерация

Наименьшей является оценка $\Delta_3 = -\frac{9}{2}M - \frac{7}{2}$. Поэтому переменная x_3 должна быть введена в число базисных. Так как в столбце A_3 только один элемент $x_{13} > 0$, то из числа базисных переменных выводится y_1 , а $x_{13} = 5$ — направляющий элемент.

Таблица 7.6

C			3	-1	2	1	-M	-M	-M
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
-M	y_1	10	2	-1	4	1	1	0	0
-M	y_2	8	-3	2	1	-2	0	1	0
-M	y_3	4	4	-1	-2	0	0	0	1
	Δ	-22M +0	-3M -3	0M +1	-3M -2	-M -1	0	0	0

Таблица 7.7

C			3	-1	2	1	-M	-M	-M
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
-M	y_1	8	0	$-\frac{1}{2}$	5	1	1	0	$-\frac{1}{2}$
-M	y_2	11	0	$\frac{5}{4}$	$-\frac{1}{2}$	-2	0	1	$\frac{3}{4}$
3	x_1	1	1	$-\frac{1}{4}$	$-\frac{1}{2}$	0	0	0	$\frac{1}{4}$
	Δ	$-19M$ +3	0 +0	$-\frac{3}{4}M$ + $\frac{1}{4}$	$-\frac{9}{2}M$ $-\frac{7}{2}$	$1M$ -1	0 0	0 0	$\frac{3}{4}M$ + $\frac{3}{4}$

Таблица 7.8

C			3	-1	2	1	-M	-M	-M
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
2	x_3	8	0	$-\frac{1}{10}$	1	$\frac{1}{5}$	$\frac{1}{5}$	0	$-\frac{1}{10}$
-M	y_2	$\frac{59}{5}$	0	$\frac{6}{5}$	0	$-\frac{19}{10}$	$\frac{1}{10}$	1	$\frac{7}{10}$
3	x_1	$-\frac{9}{5}$	1	$-\frac{3}{10}$	0	$\frac{1}{10}$	$\frac{1}{10}$	0	$\frac{1}{5}$
	Δ	$-\frac{59}{5}M$ + $\frac{43}{5}$	0 0	$-\frac{6}{5}M$ $-\frac{1}{10}$	0 0	$\frac{19}{10}M$ $-\frac{3}{10}$	$\frac{9}{10}M$ $-\frac{7}{10}$	0 0	$\frac{3}{10}M$ + $\frac{2}{5}$

Таблица 7.9

C			3	-1	2	1	-M	-M	-M
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
2	x_3	$\frac{31}{12}$	0	0	1	$\frac{1}{24}$	$\frac{5}{24}$	$\frac{1}{12}$	$-\frac{1}{24}$
-1	x_2	$\frac{59}{6}$	0	1	0	$-\frac{19}{12}$	$\frac{1}{12}$	$\frac{5}{6}$	$\frac{7}{12}$
3	x_1	$\frac{19}{4}$	1	0	0	$-\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
	Δ	0M $\frac{115}{2}$	0M 0	0M 0	0M 0	0M $-\frac{11}{24}$	1 + $\frac{17}{24}$	1 + $\frac{1}{12}$	1 + $\frac{11}{24}$

Таблица 7.10

C			3	-1	2	1
	B_K	A_0	A_1	A_2	A_3	A_4
1	x_4	62	0	0	24	1
-1	x_2	108	0	1	38	0
3	x_1	28	1	0	9	0
	Δ	38	0	0	11	0

Проделав симплексные преобразования, получим табл. 7.8.

III итерация

В строке Δ только один отрицательный элемент $\Delta_2 = -\frac{6}{5}M - \frac{1}{10}$, а в столбце A_2 только один положительный элемент $x_{22} = \frac{6}{5}$. Проделав симплексные преобразования с этим элементом, получим таблицу, в базисе которой уже не содержатся искусственные переменные. Поэтому столбцы A_5 , A_6 и A_7 , соответствующие искусственным векторам, можно не рассматривать (табл. 7.9).

IV итерация

Поскольку в таблице имеется оценка $\Delta_4 = -\frac{11}{24} < 0$, то решение $\left(\frac{19}{4}, \frac{59}{6}, \frac{31}{12}, 0\right)$ не оптимальное. В число базисных переменных вводится x_4 , а выводится переменная x_3 . Проделав симплексные преобразования, получим табл. 7.10.

Все элементы строки Δ положительны. Поэтому получено оптимальное решение $\bar{X} = (28, 108, 0, 62)$, которому соответствует оптимальное значение целевой функции $a_{00} = 38$.

Двойственность в линейном программировании

Каждой исходной задаче линейного программирования соответствует другая, вполне определенная задача линейного программирования, называемая *двойственной*. Связь между этими задачами такая, что при решении одной из них фактически решается и другая.

Если $\bar{B}$ — оптимальный базис одной из задач и $\bar{C}_B$ — вектор коэффициентов ее целевой функции, то оптимальное решение двойственной задачи равно $\bar{C}_B \bar{B}^{-1}$. Поэтому, если в двойственной задаче число ограничений меньше, чем в исходной, то эффективность вычислительного процесса при ее решении будет выше.

Понятие двойственности используется при анализе моделей задач на чувствительность.

Пусть задана исходная задача в канонической форме

$$\max x_0 = \sum_{j=1}^n c_j x_j \quad (7.45)$$

при ограничениях

$$\sum_{j=1}^n a_{ij}x_j = a_{i0}, \quad i = \overline{1, m}; \quad (7.46)$$

$$x_j \geq 0, \quad j = \overline{1, n}. \quad (7.47)$$

Ей соответствует двойственная задача

$$\min y_0 = \sum_{i=1}^m a_{i0}y_i \quad (7.48)$$

при ограничениях

$$\sum_{i=1}^m a_{ij}y_i \geq c_j, \quad j = \overline{1, n}; \quad (7.49)$$

$$y_i \text{ — произвольные для } i = \overline{1, m}. \quad (7.50)$$

Если исходная задача является задачей минимизации, то двойственная задача будет иметь вид

$$\max y_0 = \sum_{i=1}^m a_{i0}y_i \quad (7.51)$$

при ограничениях

$$\sum_{i=1}^m a_{ij}y_i \leq c_j, \quad j = \overline{1, n}; \quad (7.52)$$

$$y_i \text{ — произвольные для } i = \overline{1, m}. \quad (7.53)$$

Частным случаем рассмотренных задач являются симметричные двойственные задачи:

Исходная задача

Двойственная задача

$$\max x_0 = \sum_{j=1}^n c_j x_j;$$

$$\min y_0 = \sum_{i=1}^m a_{i0} y_i;$$

$$\sum_{j=1}^n a_{ij} x_j \leq a_{i0}, \quad i = \overline{1, m};$$

$$\sum_{i=1}^m a_{ij} y_i \geq c_j, \quad j = \overline{1, n};$$

$$x_j \geq 0, \quad j = \overline{1, n}.$$

$$y_i \geq 0, \quad i = \overline{1, m}.$$

Правила построения двойственной задачи:

1. Каждому i -у ограничению исходной задачи соответствует переменная y_i двойственной задачи и наоборот, каждому j -у ограничению двойственной задачи соответствует переменная x_j исходной задачи.

2. Матрица A ограничений исходной задачи и матрица A^T ограничений двойственной задачи взаимно транспонированы: строка коэффициентов a_{ij} в j -ом ограничении двойственной задачи является столбцом коэффициентов при x_j в ограничениях исходной задачи и наоборот.

3. Свободные члены ограничений одной из задач являются коэффициентами при соответствующих переменных в целевой функции другой задачи; максимизация меняется на минимизацию и наоборот.

4. Знаки неравенств ограничений в двойственной симметричной задаче меняются на противоположные знакам неравенств ограничений исходной задачи. Если исходная задача является задачей максимизации, двойственная к ней задача будет иметь ограничения-неравенства со знаками « $\geq$ »; если исходная задача — задача минимизации, двойственная к ней задача будет иметь ограничения-неравенства со знаками « $\leq$ ».

5. Каждому i -у ограничению-неравенству исходной задачи в двойственной задаче соответствует условие неотрицательности $y_i \geq 0$, а равенству — переменная без ограничений на знак («произвольная»).

Пример 1. Построить двойственную задачу к исходной, заданной в канонической форме:

$$\max x_0 = -3x_1 + 5x_2 + x_3 - x_4$$

при ограничениях

$$3x_1 + 8x_2 + x_3 + x_4 = 50; \quad 5x_1 - 4x_2 - x_3 + x_4 = 14;$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

Двойственная задача

$$\min y_0 = 50y_1 + 14y_2$$

при ограничениях

$$3y_1 + 5y_2 \geq -3; \quad 8y_1 - 4y_2 \geq 5; \quad y_1 - y_2 \geq 1; \quad y_1 + y_2 \geq -1,$$

где y_1 и y_2 — произвольные.

Пример 2. Построить двойственную задачу к исходной, заданной в форме общей модели:

$$\min x_0 = x_1 - 2x_2 + x_3 - x_4 + x_5$$

при ограничениях

$$x_1 - 2x_2 + x_3 + 3x_4 - 2x_5 = 6; \quad 2x_1 + 3x_2 - 2x_3 - x_4 + x_5 \leq 4;$$

$$x_1 + 3x_3 - 4x_5 \geq 8; \quad x_j \geq 0, \quad j = \overline{1, 5}.$$

Двойственная задача

$$\max y_0 = 6y_1 - 4y_2 + 8y_3$$

при ограничениях

$$y_1 - 2y_2 + y_3 \leq 1; \quad -2y_1 - 3y_2 + 0y_3 \leq -2; \quad y_1 - 2y_2 + 3y_3 \leq 1;$$

$$3y_1 + y_2 + 0y_3 \leq -1; \quad -2y_1 - y_2 - 4y_3 \leq 1,$$

где $y_2 \geq 0$; $y_3 \geq 0$; y_1 — произвольный.

Рассмотрим двойственную задачу к исходной о составлении бетона, приведенной выше.

Модель исходной задачи:

$$\min \sum_{j=1}^4 c_j x_j$$

при ограничениях

$$\sum_{j=1}^4 a_{ij} x_j \geq b_i, \quad i = \overline{1, 3};$$

$$x_j \geq 0, \quad j = \overline{1, 4},$$

где x_j — искомые количества сырья j -го вида ($j = \overline{1, 4}$); c_j — стоимость единицы сырья j -го вида; b_i — количество компонента i -го вида сырья, не меньше которого должно содержаться в бетоне ($i = \overline{1, 3}$); a_{ij} — количество единиц компонента i -го вида в сырье j -го вида.

Требуется определить количество сырья каждого вида, которое необходимо для составления бетона таким образом, чтобы в нем содержалось количество компонентов не менее заданного и стоимость бетона при этом была минимальной.

Модель двойственной задачи:

$$\max \sum_{i=1}^3 b_i y_i$$

при ограничениях

$$\sum_{i=1}^3 a_{ij} y_i \leq c_j, \quad j = \overline{1, 4};$$

$$y_i \geq 0, \quad i = \overline{1, 3}.$$

Экономическая интерпретация полученной двойственной задачи:
 y_i — стоимость (ценность) единицы i -го вида компонента бетона ($i = \overline{1, 3}$);

$\sum_{i=1}^3 a_{ij} y_i$ — стоимость (ценность) всех трех компонентов в сырье j -го вида. Эта стоимость не должна превышать стоимости единицы сырья j -го вида:

$$\sum_{i=1}^3 a_{ij} y_i \leq c_j.$$

Стоимость всех компонентов сырья всех видов, вошедших в бетон, т. е. ценность бетона, равна $\sum_{i=1}^3 b_i y_i$.

Таким образом, необходимо определить стоимость (ценность) единицы каждого i -го вида компонента бетона y_i , $i = \overline{1, 3}$, так, чтобы стоимость всех компонентов сырья j -го вида ($j = \overline{1, 4}$), содержащегося в единице сырья, не превышала стоимости единицы сырья этого вида и ценность бетона была максимальной. Часто y_i называют *теневыми стоимостями*.

Теоремы двойственности. Теоремы двойственности утверждают следующие положения:

1. Если исходная и двойственная к ней задачи имеют допустимые решения, то существуют оптимальные решения и исходной, и двойственной задач, а значения целевых функций при оптимальных решениях совпадают:

$$\sum_{j=1}^n c_j x_j^0 = \sum_{i=1}^m a_{i0} y_i^0,$$

где x_j^0 и y_i^0 — компоненты векторов оптимальных решений: $\bar{X}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ и $\bar{Y}^0 = (y_1^0, y_2^0, \dots, y_m^0)$ соответственно.

2. Если исходная (двойственная) задача имеет оптимальное решение, для которого значение целевой функции ограничено, то соответствующая ей двойственная (исходная) задача имеет оптимальное решение при том же значении целевой функций.

3. Если исходная (двойственная) задача линейного программирования имеет неограниченное оптимальное решение, то соответствующая двойственная (исходная) задача не имеет ни одного допустимого решения.

4. Любое допустимое решение задачи линейного программирования накладывает ограничение на оптимальное значение целевой функции соответствующей двойственной задачи.

Приведенные утверждения имеют большое практическое значение, так как допустимые решения исходной и двойственной к ней задач позволяют оценить оптимальное значение целевой функции задачи, т. е. указанные решения определяют интервал, в котором находится оптимальное значение целевой функции. Так же можно проверить на оптимальность любое допустимое базисное решение исходной задачи. Если существует допустимое решение двойственной задачи, для которого значение целевой функции совпадает со значением целевой функции исходной задачи, то решения обеих задач являются оптимальными.

Праведливость приведенных утверждений вытекает из следующего.

Умножим каждое i -е ограничение исходной задачи на соответствующую компоненту $y_i \geq 0$ допустимого решения двойственной задачи, а каждое j -е ограничение двойственной задачи на соответствующую компоненту $x_j \geq 0$ допустимого решения исходной задачи. Сложив отдельно правые и соответственно левые части полученных соотношений, получим:

$$\sum_{i=1}^m y_i \left(\sum_{j=1}^n a_{ij} x_j \right) \leq \sum_{i=1}^m a_{i0} y_i;$$

$$\sum_{j=1}^n x_j \left(\sum_{i=1}^m a_{ij} y_i \right) \geq \sum_{j=1}^n c_j x_j.$$

Выражения в левых частях полученных соотношений равны, поэтому

$$\sum_{j=1}^n c_j x_j \leq \sum_{i=1}^m a_{i0} y_i.$$

Следовательно, значение целевой функции, соответствующее некоторому допустимому решению (включая оптимальное) исходной задачи, не является независимым от значения целевой функции для любого допустимого решения соответствующей двойственной задачи.

Теорема о дополнительной нежесткости. Решения исходной и двойственной к ней задач являются оптимальными тогда и только тогда, когда выполняются соотношения:

$$y_i^0 \left(\sum_{j=1}^n a_{ij} x_j^0 - a_{i0} \right) = 0, \quad i = \overline{1, m};$$

$$x_j^0 \left(\sum_{i=1}^m a_{ij} y_i^0 - c_j \right) = 0, \quad j = \overline{1, n},$$

где x_j^0 и y_i^0 — компоненты векторов оптимальных решений соответственно исходной и двойственной к ней задачи:

$$\bar{X}^0 = (x_1^0, \dots, x_n^0); \quad \bar{Y}^0 = (y_1^0, \dots, y_m^0).$$

Таким образом, если при подстановке компонентов оптимального решения в систему ограничений исходной задачи i -е ограничение обращается в неравенство, то i -я компонента оптимального решения двойственной задачи равна нулю. Если i -й компонент оптимального решения двойственной задачи положителен, то i -е ограничение исходной задачи удовлетворяется ее оптимальным решением как строгое равенство.

Аналогично, если при подстановке компонентов оптимального решения в систему ограничений двойственной задачи j -е ограничение обращается в неравенство, то j -й компонент оптимального решения исходной задачи равен нулю. Если j -й компонент оптимального решения исходной задачи положителен, то j -е ограничение двойственной задачи удовлетворяется ее оптимальным решением как строгое равенство.

Двойственный симплекс-метод. В основе двойственного симплекс-метода лежит идея построения на каждой итерации, за исключением последней, решения, недопустимого вследствие невыполнения условий неотрицательности переменных; при этом соответствующее решение двойственной задачи при каждой итерации является допустимым. Этот метод в ряде случаев облегчает выбор исходного базиса без использования свободных (искусственных) переменных, помогает выполнить некоторые виды анализа модели на чувствительность и заключается в последовательном переходе от одного n -мерного вектора $\bar{X}$ к другому так, что через конечное число переходов либо получается оптимальное решение исходной задачи, либо устанавливается ее неразрешимость.

Процесс решения строится так, что как только очередной вектор $\bar{X}$ получится неотрицательным (допустимым базисным решением), он будет ее оптимальным решением.

Каждому вектору $\bar{X}$ соответствует допустимое решение двойственной задачи, причем каждому следующему вектору — лучшее, чем предыдущее.

Оптимальному решению исходной задачи соответствует оптимальное решение двойственной задачи. Таким образом, двойственный симплекс-метод — это применение обычного симплекс-метода к двойственной задаче, дополненное построением на каждой итерации n -мерного вектора $\bar{X}$, являющегося почти допустимым опорным решением исходной задачи.

Двойственным симплекс-методом целесообразно пользоваться при решении задач линейного программирования в следующих случаях:

- 1) начальное опорное допустимое базисное решение не очевидно, для его определения необходимо использовать специальные методы, например метод искусственных переменных, метод Данцига;
- 2) после получения решения задачи линейного программирования

ния необходимо добавить к исходным ограничениям какие-то дополнительные ограничения (например, при решении задач целочисленного программирования).

Пусть исходная задача линейного программирования задана в канонической форме:

$$\max x_0 = \sum_{j=1}^n c_j x_j; \quad (7.54)$$

$$\sum_{j=1}^n a_{ij} x_j = a_{i0}, \quad i = \overline{1, m}; \quad (7.55)$$

$$x_j \geq 0, \quad j = \overline{1, n}. \quad (7.56)$$

Ей соответствует двойственная задача:

$$\min y_0 = \sum_{i=1}^m a_{i0} y_i; \quad (7.57)$$

$$\sum_{i=1}^m a_{ij} y_i \geq c_j, \quad j = \overline{1, n}; \quad (7.58)$$

$$y_i — произвольные, \quad i = \overline{1, m}. \quad (7.59)$$

Пусть $\bar{Y} = (y_1, y_2, \dots, y_m)$ — допустимое базисное (опорное) решение задачи (7.57) — (7.59).

Введем ряд определений.

Базисом опорного плана (решения) $\bar{Y} = (y_1, y_2, \dots, y_m)$, или *сопряженным базисом*, называется произвольная система m линейно независимых векторов $\{\bar{A}_{k_1}, \dots, \bar{A}_{k_m}\}$ ограничений задачи (7.54) — (7.56), для которых m ограничений (7.58) выполняются как строге равенства.

Вектор $\bar{X} = (x_1, x_2, \dots, x_n)$, компоненты которого равны коэффициентам разложения x_{i0} вектора $\bar{A}_0$ по векторам сопряженного базиса $\{\bar{A}_{k_1}, \dots, \bar{A}_{k_m}\}$, называется *псевдопланом исходной задачи* (7.54) — (7.56).

Псевдоплан $\bar{X}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ будет оптимальным решением исходной задачи, если среди его базисных компонентов нет отрицательных, т. е. $x_{i0}^0 \geq 0, i = \overline{1, m}$. План $\bar{Y}^0 = (y_1^0, y_2^0, \dots, y_m^0)$, определяющий псевдоплан $\bar{X}^0$, является решением двойственной задачи (7.57) — (7.59).

Если среди базисных компонентов псевдоплана $\bar{X} = (x_1, x_2, \dots, x_n)$ имеются отрицательные компоненты, и среди них есть хотя бы одна отрицательная компонента $x_{i0} < 0$, для которой $x_{ij} \geq 0, j = \overline{1, n}$, то целевая функция исходной задачи (7.54) и (7.56) неограничена снизу на множестве ее планов, т. е. задача (7.54) и (7.56) — неразрешима; если для каждой компоненты $x_{i0} < 0$ существует хотя бы один $x_{ij} < 0, j = \overline{1, n}$, то данный псевдоплан $\bar{X} = (x_1, x_2, \dots, x_n)$ можно улучшить.

Для улучшения псевдоплана из базиса выводится вектор $\bar{A}_{k_r}$, причем, если имеется несколько $x_{rj} < 0$, то из базиса можно вывести любой из векторов $\bar{A}_{k_r}$, например тот, для которого x_{rj} имеет минимальное значение.

Вводится в базис вектор $\bar{A}_l$, номер которого определяется из соотношения

$$-\frac{\Delta_l}{x_{rl}} = \min \left\{ -\frac{\Delta_j}{x_{rj}} \mid x_{rj} < 0 \right\}, \quad (7.60)$$

Приведем алгоритм двойственного симплекс-метода.

1. Представить исходную задачу линейного программирования в канонической форме (7.54)—(7.56).

2. Записать двойственную задачу (7.57)—(7.59).

3. Выбрать сопряженный базис по ограничениям (7.58)—(7.59).

4. Определить коэффициенты разложения небазисных векторов по векторам сопряженного базиса.

5. Записать псевдоплан $\bar{X} = (x_1, \dots, x_n)$.

6. Проверить, имеются ли среди компонентов псевдоплана $\bar{X} = (x_1, \dots, x_n)$ отрицательные ($x_j < 0$). Если не имеются, то план $\bar{X} = (x_1, \dots, x_n)$ — оптимальный план исходной задачи (7.54)—(7.56), в противном случае выполняют условие п. 7.

7. Проверить в строках псевдоплана, соответствующих его отрицательным компонентам ($x_{r0} < 0$), все ли элементы $x_{rj} \geq 0$, $j = \overline{1, n}$. Если да, то задача (7.54)—(7.56) — неразрешима. Если нет, выполняют условие п. 8.

8. Выбрать направляющую строку, т. е. определить вектор, выводимый из базиса. Если $x_{rj} < 0$ при нескольких r , то из базиса можно вывести любой из векторов $\bar{A}_{h_r}$.

9. Определить направляющий столбец, т. е. вектор, вводимый в базис $\bar{A}_l$. Номер этого столбца l определяется из соотношения

$$\Theta_{\min} = \min \left\{ -\frac{\Delta_j}{x_{rj}} \mid x_{rj} < 0 \right\} = -\frac{\Delta_l}{x_{rl}}.$$

10. Выполнить симплексные преобразования и условие п. 6.

Пример. Рассмотрим решение задачи о составлении бетона двойственным симплекс-методом при следующей исходной информации:

стоимости единицы каждого вида сырья:

$$c_1 = 2, \quad c_2 = 4, \quad c_3 = 3, \quad c_4 = 5;$$

необходимое количество компонентов бетона:

$$b_1 = 600, \quad b_2 = 250, \quad b_3 = 150;$$

содержание i -й компоненты в единице j -го вида сырья a_{ij} :

$$\begin{aligned} a_{11} = 6; \quad a_{12} = 6; \quad a_{13} = 0; \quad a_{14} = 4; \quad a_{21} = 0; \quad a_{22} = 2; \quad a_{23} = 5; \\ a_{24} = 6; \quad a_{31} = 4; \quad a_{32} = 2; \quad a_{33} = 5; \quad a_{34} = 0. \end{aligned}$$

Математическая модель задачи (7.12)—(7.13):

минимизировать

$$(2x_1 + 4x_2 + 3x_3 + 5x_4)$$

при ограничениях

$$\begin{aligned} 6x_1 + 6x_2 + 0x_3 + 4x_4 &\geq 600; & 4x_1 + 2x_2 + 5x_3 + 0x_4 &\geq 150; \\ 0x_1 + 2x_2 + 5x_3 + 6x_4 &\geq 250; & x_1 \geq 0, \quad x_2 \geq 0; \quad x_3 \geq 0, \quad x_4 &\geq 0. \end{aligned}$$

Решение. Запишем модель задачи в канонической форме: максимизировать

$$(-2x_1 - 4x_2 - 3x_3 - 5x_4)$$

Таблица 7.11

c			-2	-4	-3	-5	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
+0	x_5	-600	-6	-6	0	-4	1	0	0
0	x_6	-250	0	-2	-5	-6	0	1	0
0	x_7	-150	-4	-2	-5	0	0	0	1
	Δ	0	2	4	3	5	0	0	0
	Θ		$\frac{2}{6}$	$\frac{4}{6}$		$\frac{5}{4}$			

при ограничениях

$$6x_1 + 6x_2 + 0x_3 + 4x_4 - 1x_5 = 600;$$

$$0x_1 + 2x_2 + 5x_3 + 6x_4 - 1x_6 = 250;$$

$$4x_1 + 2x_2 + 5x_3 + 0x_4 - 1x_7 = 150;$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0, x_7 \geq 0.$$

Поскольку модель задачи не имеет единичного базиса, т. е. нельзя просто построить начальное допустимое решение, используем двойственный симплекс-метод.

Для выбора сопряженного базиса построим двойственную задачу: минимизировать

$$(-600y_1 - 250y_2 - 150y_3)$$

при ограничениях

$$\begin{array}{l|l|l} 6y_1 + 0y_2 + 4y_3 \leq 2 & (\bar{A}_1); & y_1 \geq 0 & (\bar{A}_5); \\ 6y_1 + 2y_2 + 2y_3 \leq 4 & (\bar{A}_2); & y_2 \geq 0 & (\bar{A}_6); \\ 0y_1 + 5y_2 + 5y_3 \leq 3 & (\bar{A}_3); & y_3 \geq 0 & (\bar{A}_7); \\ 4y_1 + 6y_2 + 0y_3 \leq 5 & (\bar{A}_4); & & \end{array}$$

Сопряженным является базис $(\bar{A}_5, \bar{A}_6, \bar{A}_7)$, так как $y_1 = 0, y_2 = 0, y_3 = 0$, т. е. удовлетворяют всем ограничениям двойственной задачи.

Псевдоплан будет иметь вид

$$\bar{X} = (0, 0, 0, 0 - 600, -250, -150).$$

Заполним симплекс-таблицу 7.11.

I итерация

Поскольку $x_5 = -600$ — наименьший отрицательный компонент псевдоплана, то x_5 выводится из числа базисных переменных. Переменная x_1 вводится в число базисных переменных, так как $\Theta_{\min} = \Theta_1 = 2/6$.

Выполнив симплекс-преобразования с направляющим элементом x_{11} , получим симплекс-таблицу 7.12.

II итерация

В псевдоплане только одна отрицательная компонента $x_6 = -250$. Поэтому переменную x_6 необходимо вывести из псевдоплана, а ввести в псевдоплан переменную x_3 , так как $\Theta_{\min} = \Theta_3 = 3/5$.

Прделаав симплексные преобразования с направляющим элементом x_{23} , получим симплекс-таблицу 7.13.

Таблица 7.12

C			-2	-4	-3	-5	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
-2	x_1	100	1	1	0	$\frac{2}{3}$	$-\frac{1}{6}$	0	0
0	x_6	-250	0	-2	-5	-6	0	1	0
0	x_7	250	0	2	-5	$\frac{8}{3}$	$-\frac{2}{3}$	0	1
	Δ	-200	0	2	3	$\frac{11}{3}$	$\frac{1}{3}$	0	0
	Θ			$\frac{2}{2}$	$\frac{3}{5}$	$\frac{11}{36}$			(

Таблица 7.13

C			-2	-4	-3	-5	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
-2	x_1	100	1	1	0	$\frac{2}{3}$	$-\frac{1}{6}$	0	0
-3	x_3	50	0	$\frac{2}{5}$	1	$\frac{6}{5}$	0	$-\frac{1}{5}$	0
0	x_7	500	0	4	0	$\frac{26}{3}$	$-\frac{2}{3}$	-1	1
	Δ	-350	0	$\frac{4}{5}$	0	$\frac{109}{15}$	$\frac{1}{3}$	$\frac{3}{5}$	0
	Θ								

Поскольку все компоненты псевдоплана неотрицательны, то получено оптимальное решение задачи:

$$\bar{X}^0 = (100, 0, 50, 0), \quad a_{00} = -350,$$

значение целевой функции a_{00} соответствует задаче максимизации.

Таким образом, оптимальное решение задачи о составлении бетона следующее:

для составления бетона необходимо взять 100 ед. сырья первого вида ($x_1 = 100$) и 50 ед. сырья третьего вида ($x_3 = 50$), при этом стоимость бетона равна 350 ед.

Анализ модели на чувствительность. Специалиста в области исследования операций интересует не только полученное статическое оптимальное решение задачи, но также и те изменения оптимального решения, которые связаны с изменением в соответствующих моделях компонентов векторов $\bar{C}$ (коэффициентов целевой функции) и $\bar{A}_0$ (правых частей ограничений), а также компонентов матрицы $\bar{A}$ (коэффициентов при неизвестных в ограничениях), т. е. интерес представляет динамическая информация о решении, свя-

занная с изменением наличных ресурсов различных видов, стоимостей, различных затрат и т. п.

В большинстве случаев представляют интерес ответы на такие вопросы: в каком интервале можно менять входные параметры без существенного отклонения от найденного оптимального решения и без значительного нарушения структуры базиса, формирующего оптимальное решение? Изменится ли полученное оптимальное решение, и если да, то каково новое оптимальное решение?

Исследование, позволяющее ответить на эти вопросы, называется *анализ модели на чувствительность* (или анализ модели при известном оптимальном решении).

Численные данные заключительной симплекс-таблицы позволяют ответить на такие вопросы: остается ли решение оптимальным, если уменьшить удельный вклад в прибыль одной из базисных переменных? К каким последствиям приведет сокращение объема ресурсов? Что произойдет, если ввести в рассмотрение новую управляемую переменную?

Теория двойственности дает возможность проанализировать модель на чувствительность.

1. Пусть изменяются компоненты вектора $\bar{A}_0 = (a_{i0}, \dots, a_{m0})$, т. е. изменяются ресурсы.

Как эти изменения повлияют на оптимальное значение целевой функции? В каких случаях найденное оптимальное решение $\bar{X}^0$ не изменится?

Допустим величина a_{i0} получила некоторое приращение Δa_{i0} . Соответственно изменится оптимальный план $\bar{X}^0 = \bar{X}^0(a_{i0} + \Delta a_{i0})$ и значение функции цели $f_{\max}(\bar{X}^0(a_{i0} + \Delta a_{i0}))$.

Пусть Δa_{i0} таковы, что решение $\bar{X}^0 = \bar{X}^0(a_{i0} + \Delta a_{i0})$ остается допустимым, т. е. $\bar{X}^0 \geq 0$.

Найдем отношение

$$\frac{\Delta f_{\max}^0(a_{i0})}{\Delta a_{i0}} = \frac{f_{\max}^0(\bar{X}^0(a_{i0} + \Delta a_{i0})) - f_{\max}^0(\bar{X}^0(a_{i0}))}{\Delta a_{i0}}. \quad (7.61)$$

При $\Delta a_{i0} \rightarrow 0$, получим

$$\lim_{\Delta a_{i0} \rightarrow 0} \frac{\Delta f_{\max}^0(a_{i0})}{\Delta a_{i0}} = \frac{\partial f_{\max}^0(a_{i0})}{\partial a_{i0}}. \quad (7.62)$$

На основании теоремы двойственности имеем:

$$f_{\max}^0(\bar{X}^0) = \sum_{j=1}^n c_j x_j^0 = \sum_{i=1}^m a_{i0} y_i^0. \quad (7.63)$$

Подставляя (7.63) в (7.62), получим

$$\frac{\partial f_{\max}^0(a_{i0})}{\partial a_{i0}} = y_i^0, i = \overline{1, m}. \quad (7.64)$$

Таким образом, управляемые переменные оптимального решения двойственной задачи $\bar{Y}^0 = (y_1^0, \dots, y_m^0)$ определяют вклад каж-

дого i -го ресурса в доход $f_{\max}^{0i}$ при оптимальном решении $\bar{X}^0$. Эта величина равна допустимому доходу при увеличении ресурса i -го вида, т. е. правой части ограничения a_{i0} , на единицу и условии, что ресурсы используются оптимальным образом. Управляемые переменные оптимального решения $\bar{Y}^0$ являются показателями важности соответствующих ресурсов для исследуемой системы управления. Чем больше значение y_i^0 , тем существеннее вклад i -го ресурса в максимальный доход $f_{\max}^{0i}$ и тем выгоднее его увеличение.

Если $y_i^0 = 0$, то i -й ресурс не является существенным ограничением для системы, так как по условию дополняющей нежесткости

$$a_{i0} - \sum_{j=1}^n a_{ij}x_j^0 > 0.$$

Поэтому величины y_i^0 называют «скрытыми» доходами, или маргинальными оценками управляемой системы.

2. Пусть изменяются коэффициенты целевой функции c_j , $j = \overline{1, n}$. Рассмотрим условия, при которых найденное оптимальное решение не изменится. Допустим новое значение коэффициента в целевой функции отличается от старого на δ_l , т. е.

$$c_l^1 = c_l + \delta_l.$$

Пусть I — множество индексов базисных переменных оптимального решения $\bar{X}^0$. Найдем оценки (симплекс-разности) Δ_j^1 после изменения c_l на δ_l для двух случаев: c_l — коэффициент при небазисной переменной x_l , $l \notin I$, и c_l — коэффициент при базисной переменной x_l , $l \in I$.

Если $l \notin I$, то

$$\Delta_j^1 = \Delta_j \text{ для всех } j \neq l; \quad (7.65)$$

$$\Delta_j^1 = \sum_{i \in I} c_i a_{il} - (c_l + \delta_l) \text{ для } j = l. \quad (7.66)$$

Если $l \in I$, то

$$\Delta_j^1 = \sum_{i \in I} c_i^1 a_{ij} - c_j = \sum_{i \in I} c_i a_{ij} + \delta_l a_{lj} - c_j \text{ для } j \notin I. \quad (7.67)$$

Таким образом, для сохранения оптимальности прежнего решения при изменении c_l необходимо и достаточно сохранение знаков оценок Δ_j^1 для всех небазисных переменных. Поэтому из условия $\Delta_j^1 \geq 0$ (задача максимизации) можно определить допустимые изменения δ_l коэффициента c_l , при котором сохраняется прежнее оптимальное решение.

Аналогично можно получить допустимые изменения для всех или нескольких коэффициентов целевой функции. В этом случае новые оценки Δ_j будут функциями нескольких параметров ($\delta_1, \delta_2, \dots, \delta_r$). Решая совместно систему

$$\Delta_j(\delta_1, \delta_2, \dots, \delta_r) \geq 0, \quad j \in I, \quad (7.68)$$

получим искомые допустимые изменения коэффициентов целевой функции.

Рассмотрим общий случай, предполагая, что параметры модели c_j , a_{ij} , a_{i0} являются функциями некоторого параметра t .

Изменение целевой функции в окрестности оптимального решения имеет вид

$$\begin{aligned} \frac{df^0}{dt} = & - \sum_{i=1}^m \sum_{j=1}^n \frac{\partial f^0}{\partial a_{i0}} \frac{\partial f^0}{\partial c_j} \frac{da_{ij}}{dt} + \\ & + \sum_{j=1}^n \frac{\partial f^0}{\partial c_j} \frac{dc_j}{dt} + \sum_{i=1}^m \frac{\partial f^0}{\partial a_{i0}} \frac{da_{i0}}{dt}. \end{aligned} \quad (7.69)$$

Если принять, что $t = a_{ij}$, то

$$\frac{\partial f^0}{\partial a_{ij}} = - \frac{\partial f^0}{\partial a_{i0}} \frac{\partial f^0}{\partial c_j} = -y_i^0 x_j^0. \quad (7.70)$$

где y_i^0 и x_j^0 — управляемые переменные соответственно оптимального решения $\bar{Y}^0$ двойственной задачи и оптимального решения $\bar{X}^0$ исходной задачи.

Условие (7.70) показывает, что скорость изменения оптимального значения f^0 прямо пропорциональна произведению оптимальных решений исходной и двойственной к ней задач. Если y_i^0 и x_j^0 неотрицательны, то положительное приращение a_{ij} вызовет уменьшение целевой функции.

Таким образом, имея оптимальное решение исходной задачи $\bar{X}^0 = (x_1^0, \dots, x_n^0)$ и оптимальное решение двойственной к ней задачи $\bar{Y}^0 = (y_1^0, \dots, y_m^0)$, можно исследовать модель задачи на чувствительность. При этом надо помнить, что решение двойственной задачи получается из последней симплекс-таблицы с оптимальным решением исходной задачи по правилу

$$\bar{Y}^0 = \bar{C}_B \bar{B}^{-1},$$

где $\bar{C}_B$ — вектор коэффициентов при базисных переменных целевой функции исходной задачи; $\bar{B}$ — оптимальный базис исходной задачи.

Поскольку значение целевой функции f^0 зависит от управляемых переменных, которые определяются в зависимости от коэффициентов ограничений, то

$$\frac{\partial f^0}{\partial a_{ij}} = \frac{\partial f^0}{\partial x_j} \frac{\partial x_j}{\partial a_{ij}} = c_j \frac{\partial x_j}{\partial a_{ij}}. \quad (7.71)$$

В выражениях (7.70) и (7.71) левые части равны, поэтому при оптимальных значениях управляемых переменных x_j^0 и y_i^0 изменение x_j , y_i в окрестности экстремальной точки определяется в зависимости от изменения коэффициентов ограничений a_{ij} следующим образом:

$$\left. \frac{\partial x_j}{\partial a_{ij}} \right|_{x_i^0} = - \frac{y_i^0 x_j^0}{c_j}; \quad \left. \frac{\partial y_i}{\partial a_{ij}} \right|_{y_i^0} = - \frac{y_i^0 x_j^0}{a_{i0}}. \quad (7.72)$$

Сети используются для решения экстремальных задач в самых различных областях: календарное планирование, планирование материально-технического снабжения, моделирование различных процессов и т. п.

Математические характеристики сетевых моделей, которыми описываются разнообразные задачи, имеют особенности, позволяющие существенно повысить эффективность процесса отыскания оптимальных решений задач, содержащих тысячи переменных и сотни ограничений. Благодаря «сетевому языку» разнообразные, совершенно непохожие, задачи допускают применение для их решения стандартных, хорошо разработанных математических методов.

Сеть является частным случаем графа. *Граф* — это совокупность точек (узлов, вершин) и линий, соединяющих эти точки. Графом описывают отношения между множествами объектов, которые обладают многими характеристиками. При исследовании задач, описываемых графами, особое значение имеют связи между отдельными точками (объектами) и порядок их соединения.

Все задачи, решаемые с помощью теории графов, делятся на две группы: 1) требуется определить, существуют ли объекты с определенными свойствами и сколько этих объектов; 2) заданы некоторые свойства объектов и требуется построить граф.

Математический граф G — это упорядоченная пара $G = (V, L)$, где V — множество вершин (непустое множество), L — множество ребер.

Если ребро принадлежит множеству L , то оно инцидентно вершинам (узлам), которые оно соединяет. Например, ребро l_j ($l_j \in L$) соединяет вершины v_{i_1} и v_{i_2} ($v_{i_1} \in V$, $v_{i_2} \in V$). Значит ребро l_j инцидентно вершинам v_{i_1} и v_{i_2} . Ребра и вершины являются элементами графа.

Графы делятся на ориентированные и неориентированные.

Ориентированный граф — это упорядоченная пара (V, L) , т. е. L — упорядоченное отношение на V . Ребра упорядоченного графа называются *дугами*. Каждая дуга соединяет две вершины. Вершина, из которой выходит дуга, называется *начальной*; вершина, в которую входит дуга, называется *конечной*.

Неориентированный граф полностью определяется множеством вершин и множеством ребер с указанием их инцидентности. Поэтому такой граф может быть полностью описан матрицей инцидентности.

Если граф содержит конечное число ребер, он называется *конечным*.

Граф может иметь подграфы. Граф $G' = (V', L')$ называется *подграфом* $G = (V, L)$, если $V' \subset V$, $L' \subset L$. Если ребро входит в ту же вершину, из которой выходит, то это ребро образует *петлю*.

Одной и той же вершине может быть инцидентно несколько ребер. *Валентностью* (степенью) *вершины* называется количество инцидентных ей ребер. Две вершины могут соединяться несколькими

ми ребрами (отражающими различные характеристики). Количество ребер, соединяющих две вершины, называется *кратностью* этой пары вершин.

Последовательность n ребер (дуг) $l_1, \dots, l_n$ называется *маршрутом длины n* , если существует соответствующая последовательность из $n+1$ вершин $v_0, v_1, \dots, v_n$ таких, что l_j инцидентно $(v_{j-1} \& v_j)$, $j = \overline{1, n}$. В ориентированном графе — маршрут ориентированный, в неориентированном — неориентированный.

Если начальная вершина маршрута v_0 совпадает с его конечной вершиной v_n , т. е. $v_0 = v_n$, то маршрут (ориентированный) называется *замкнутым*, если $v_0 \neq v_n$ — маршрут называется *незамкнутым*.

Если в маршруте $l_i \neq l_j$ для всех i и j , $i \neq j$, то маршрут называется *цепью*. Для ориентированного графа цепь является путем. Таким образом, множество ребер (дуг) образует цепь (путь).

Если начальная вершина графа совпадает с конечной, т. е. $v_0 = v_n$, то цепь (путь) называется *циклом* (контуром). Количество ребер наиболее длинного цикла называют *окружением графа*. В циклах и контурах каждое ребро должно содержаться только один раз, вершины могут повторяться.

Графы могут быть связными и несвязными. В *связном* графе каждая пара вершин соединяется цепью. В *несвязном* графе можно выделить конечное число подграфов, которые называются *компонентами*, или частями, графа.

Наименьшее количество вершин связного графа, удаление которых приводит к несвязному графу, определяет связность графа.

Деревом в неориентированном графе называется связный подграф, не имеющий циклов.

Если все вершины графа принадлежат дереву, то оно называется *покрывающим*. Если существует путь от вершины v_i к каждой другой вершине графа, то ориентированное дерево имеет *корень* в этой вершине v_i .

В покрывающем дереве ребра графа соответствуют *ветви*. Все остальные ребра называются *хордами*.

Сетью называется линейный граф, элементам которого поставлены в соответствие некоторые параметры.

Задачи, связанные с принятием плановых решений, описываются сетевыми моделями. Решение этих задач обычно сводится к решению одной из следующих задач: минимизации сети, определения кратчайшего маршрута, определения максимального потока.

Общая линейная экстремальная задача на сети

Пусть имеется некоторый конечный граф $G = (V, L)$, каждой вершине которого поставлено в соответствие некоторое число g_i , называемое *интенсивностью вершины*.

Вершины называются источниками, нейтральными вершинами и стоками, если $g_i > 0$, $g_i = 0$ и $g_i < 0$ соответственно, т. е.

$$V = V^+ \cup V^0 \cup V^-,$$

где

$$V^+ = \{i: g_i > 0\}, \quad V^- = \{i: g_i < 0\}, \quad V^0 = \{i: g_i = 0\}.$$

Потоком в сети называется совокупность величин, удовлетворяющих соотношениям:

$$\sum_j x_{ij} - \sum_j x_{ji} = g_i, \quad i \in V; \quad (7.73)$$

$$x_{ij} \geq 0, \quad (i, j) \in L. \quad (7.74)$$

Соотношение (7.73) физически означает, что поток, выходящий из i -й вершины, равен сумме ее интенсивности и входящего в нее потока.

Требуется найти такой поток в сети, который минимизирует целевую функцию

$$\sum_{(i,j) \in L} c_{ij} x_{ij}, \quad (7.75)$$

где c_{ij} , $(i, j) \in L$ — функция стоимостей потока.

Такой поток называется *оптимальным*.

Задача (7.73) — (7.75) является представлением задачи линейного программирования на сети и называется *задачей об оптимальном потоке* (максимальном потоке).

Классическая транспортная задача

Задача формулируется следующим образом: имеется m различных поставщиков (предприятий или пунктов отправления), которые имеют некоторый однородный продукт (изделия). Этот продукт требуется n потребителям, т. е. его нужно отправить в n пунктов назначения, связанных с поставщиками дорогами. Предполагается, что поставщик i ($i = \overline{1, m}$) может отгрузить не более a_i продукта (груза), потребитель j требует не менее b_j продукта (груза). Величины a_i , $i = \overline{1, m}$, и b_j , $j = \overline{1, n}$, на плановом периоде принимаются постоянными. Заданы затраты c_{ij} на перевозку единицы груза из пункта отправления i в пункт назначения j . Необходимо так составить план перевозок грузов в плановом периоде, чтобы их общая стоимость была минимальной.

Математическая модель такой задачи имеет вид: минимизировать

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (7.76)$$

при ограничениях

$$\sum_{j=1}^n x_{ij} \leq a_i, \quad i = \overline{1, m}; \quad (7.77)$$

$$\sum_{i=1}^m x_{ij} \geq b_j, \quad j = \overline{1, n}; \quad (7.78)$$

$$x_{ij} \geq 0, \quad i = \overline{1, m}; \quad j = \overline{1, n}. \quad (7.79)$$

Задача (7.76) — (7.79) будет иметь решение только в случае, когда

общие ресурсы поставщиков будут не меньше общего спроса потребителей:

$$\sum_{i=1}^m a_i \geq \sum_{j=1}^n b_j. \quad (7.80)$$

Для решения задачи (7.76) — (7.79) ее необходимо представить в канонической форме, т. е. ограничения (7.77), (7.78) записать в виде равенств. Это достигается, если

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j. \quad (7.81)$$

В этом случае система ограничений будет совместна.

Для выполнения условия (7.81), в случае необходимости вводят фиктивный потребитель со спросом

$$\sum_{i=1}^m a_i - \sum_{j=1}^n b_j$$

или фиктивный поставщик с ресурсом

$$\sum_{j=1}^n b_j - \sum_{i=1}^m a_i.$$

Если номер фиктивного потребителя k , то $c_{ik}=0$. В этом случае груз x_{ik} рассматривают как «фиктивную поставку». Если номер фиктивного поставщика s , то $c_{sj}=0$, $j=\overline{1, n}$.

В дальнейшем предполагают, что условие (7.81) всегда выполняется, т. е. ограничения (7.77), (7.78) являются равенствами. Таким образом, модель транспортной задачи будет иметь вид: минимизировать

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (7.82)$$

при ограничениях

$$\sum_{j=1}^n x_{ij} = a_i, \quad i = \overline{1, m}; \quad (7.83)$$

$$\sum_{i=1}^m x_{ij} = b_j, \quad j = \overline{1, n}; \quad (7.84)$$

$$x_{ij} \geq 0, \quad i = \overline{1, m}; \quad j = \overline{1, n}. \quad (7.85)$$

Физический смысл ограничений (7.83) и (7.84) следующий: все наличные ресурсы поставщиков должны быть вывезены и все потребности потребителей должны быть удовлетворены. Для заданного периода планирования требуется составить такой план перевозки груза, при котором общие транспортные затраты минимальны. Предполагается, что перевозится груз одного вида, так как задача перевозки грузов различных видов (многопродуктовая задача) сводится к рассматриваемой.

Транспортная задача (7.82) — (7.85) обладает следующими свойствами.

Для существования решения транспортной задачи необходимо выполнение условия баланса $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$.

В связи с этим одно из ограничений является линейной комбинацией всех других, т.е. ранг матрицы ограничений (7.83)—(7.84) равен $m+n-1$. Если все a_i , $i=1, m$, и b_j , $j=1, n$, представлены целыми числами, то хотя бы одно из оптимальных решений будет целочисленным.

Допустимое решение транспортной задачи (план перевозок) записывается в виде двумерной матрицы, состоящей из m строк и n столбцов:

$$\bar{X} = \|x_{ij}\|_{m \times n},$$

где x_{ij} — количество груза (продукта), перевозимого от i -го производителя j -у потребителю.

Допустимое базисное решение транспортной задачи называется ее *опорным планом*. План является опорным, если из ненулевых его перевозок нельзя составить замкнутый маршрут (цикл).

Опорность плана проверяется следующим образом. Просматривают последовательно строки, а затем столбцы плана $\bar{X}$ и вычеркивают те, которые содержат не более одной ненулевой перевозки. Весь процесс повторяют, пока после конечного числа вычеркиваний будет один из двух исходов: все строки (столбцы) вычеркнуты; получена подматрица, в каждой строке (столбце) которой содержится не менее двух ненулевых перевозок. В первом случае составить цикл из перевозок (коммуникаций) невозможно — план опорный. Во втором случае можно составить цикл из оставшихся перевозок — план не опорный.

Поставим в соответствие пунктам отправления и пунктам назначения вершины графа, а дорогам перевозок — дуги графа. Таким образом, множество вершин графа $V = \{1, 2, \dots, m, m+1, \dots, m+n\}$, а множество дуг графа L определим так, что $(i, j) \in L$, если из пункта i имеется дорога в пункт j .

Для получения транспортной сети поставим в соответствие элементам рассмотренного графа следующие параметры:

g_s — количество единиц груза, которое имеется ($g_s \geq 0$) или потребляется ($g_s < 0$) в вершине s ($s=1, \dots, m, m+1, \dots, m+n$);

y_{ij} — количество единиц груза, которое может быть перевезено из вершины (пункта) i в вершину (пункт) j , $(i, j) \in L$, т.е. пропускная способность пути (i, j) ;

c_{ij} — стоимость перевозки единицы груза из вершины i в вершину j (можно интерпретировать как длину пути).

Если x_{ij} — искомое количество груза, перевозимое из пункта i в пункт j , то оно соответствует потоку для дуги (i, j) и удовлетворяет условию

$$0 \leq x_{ij} \leq y_{ij}, \quad (i, j) \in L. \quad (7.86)$$

Условие (7.73) также должно выполняться:

$$\sum_j x_{ij} - \sum_j x_{ji} = g_i, \quad i \in V, \quad (7.87)$$

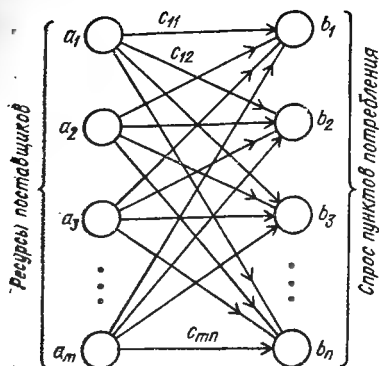


Рис. 7.2. Сеть транспортной задачи

j							
i		1	2	3	. . .	n	α_i
1		c_{11}	c_{12}	c_{13}	. . .	c_{1n}	α_1
2		c_{21}	c_{22}	c_{23}	. . .	c_{2n}	α_2
3		c_{31}	c_{32}	c_{33}	. . .	c_{3n}	α_3
.		.	.	.		.	
.		.	.	.		.	
m		c_{m1}	c_{m2}	c_{m3}	. . .	c_{mn}	α_m
b_j		b_1	b_2	b_3	. . .	b_n	

Рис. 7.3. Исходная информация к транспортной задаче

т. е. разность между количеством груза, отправленным из пункта (вершины) i , и количеством груза, пришедшим в этот же пункт (вершину), должна равняться интенсивности груза (потока).

Поскольку общие затраты на перевозку всех требуемых потребителями грузов равны

$$\sum_{(i,j) \in L} c_{ij}x_{ij}, \quad (7.88)$$

то задача (7.86) — (7.88) сводится к задаче об оптимальном потоке (7.73) — (7.75), т. е. классическая транспортная задача (7.76) — (7.78) свелась к задаче на сети при условии, что $x_{ij} = y_{ij}$, $(i, j) \in L$. Если выполняется условие $x_{ij} < y_{ij}$, $(i, j) \in L$, то модель (7.86) — (7.87) описывает транспортную задачу с ограниченными пропускными способностями.

Таким образом, сеть транспортной задачи имеет вид, изображенный на рис. 7.2.

В сети транспортной задачи имеется m источников и n стоков. Исходная информация, необходимая для решения задачи, может быть записана в виде таблицы (двумерного массива), элементами которой являются стоимости перевозок c_{ij} единицы груза от i -го поставщика к j -му потребителю; в последней строке записывается спрос b_j , $j = \overline{1, n}$, в последнем столбце — наличные ресурсы a_i , $i = \overline{1, m}$ (рис. 7.3).

Решение транспортной задачи методом потенциалов. Метод основывается на свойствах двойственности задач линейного программирования.

Двойственная задача к исходной транспортной задаче (7.82) — (7.85) описывается следующей моделью:

максимизировать

$$\sum_{i=1}^m a_i v_i + \sum_{j=1}^n b_j w_j \quad (7.89)$$

при ограничениях

$$v_i + w_j \leq c_{ij}, \quad i = \overline{1, m}; \quad j = \overline{1, n}; \quad (7.90)$$

$$v_i, w_j, \quad i = \overline{1, m}; \quad j = \overline{1, n} - \text{произвольные.} \quad (7.91)$$

Исходя из теоремы двойственности и теоремы о дополняющей нежесткости, можно утверждать, что если управляющие переменные $x_{ij}^0, i = \overline{1, m}; j = \overline{1, n}$, исходной задачи удовлетворяют ограничениям (7.83)—(7.85) и управляемые переменные v_i^0 и w_j^0 двойственной задачи удовлетворяют ограничениям (7.90), а также выполняются условия

$$x_{ij}^0(v_i^0 + w_j^0 - c_{ij}) = 0 \text{ для всех } i = \overline{1, m}; j = \overline{1, n}, \quad (7.92)$$

то управляемые переменные $x_{ij}^0, i = \overline{1, m}; j = \overline{1, n}$, являются компонентами оптимального плана перевозок.

На основании изложенного разработаны алгоритмы решения сетевых задач, основанные на общей идее, что на каждой итерации для управляемых переменных x_{ij}, v_i и w_j всегда выполняются два из следующих трех условий: допустимость исходной задачи, т. е. выполнение ограничений (7.83)—(7.85); допустимость двойственной задачи, т. е. выполнение ограничений (7.90); условия дополняющей нежесткости (7.92).

В методе потенциалов (переменные v_i и $w_j, i = \overline{1, m}; j = \overline{1, n}$, двойственной задачи (7.89)—(7.91) называются соответственно потенциалами пунктов отправления и пунктов потребления) на каждой итерации выполняются первое и третье условия. Признаком оптимальности плана перевозок является выполнение третьего условия. В рассматриваемой задаче предполагается, что ни в одном маршруте, соединяющем источник с некоторым стоком, другие источники и стоки не могут быть использованы в качестве промежуточных пунктов.

Метод потенциалов состоит в следующем. Имеется некоторый начальный поток (начальное допустимое базисное решение, начальный план), на основании компонентов которого определяются потенциалы вершин сети. Если выполняются условия $v_i^0 + w_j^0 - c_{ij} = 0$, то начальный поток (план) оптимален. Если имеется (поток) дуга, для которой не выполняется условие оптимальности, то поток необходимо изменить, т. е. улучшить план перевозок. По измененному потоку определяются новые потенциалы, улучшается поток (план) и цикл повторяется, пока не будет выполняться условие оптимальности.

Таким образом, для реализации метода потенциалов необходимо уметь составлять начальный поток (план) и в случае необходимости улучшать его.

Рассмотрим построение начального плана (потока). Существует несколько методов построения начального плана (начального допустимого решения). Наиболее распространенными являются метод «северо-западного угла» и метод «минимального элемента», в основе которого лежит метод «северо-западного угла».

При методе «северо-западного угла» определяется значение x_{11} , т. е. максимально допустимый поток для маршрута, соединяющего первый источник с первым стоком:

$$x_{11} = \min\{a_1, b_1\}.$$

Если $a_1 = b_1$, то удаляются источник 1 и сток 1 и выбирается для определения x_{22} , т. е. маршрут, соответствующий x_{22} .

Если $a_1 > b_1$, то удовлетворяется спрос стока, т. е. $x_{11} = b_1$, сток удаляется и выбирается маршрут, соответствующий x_{12} .

Если $a_1 < b_1$, то предложение источника реализовано полностью, сток не удовлетворен, т. е. $x_{11} = a_1$, удаляется источник и выбирается маршрут, соответствующий x_{21} .

Таким образом, для выбранного маршрута определяется максимально допустимый поток и удаляется либо источник, если его предложение реализовано полностью, либо сток, если его спрос удовлетворен. Если одновременно выполняются условия на предложение и спрос, то удаляются и источник и сток. Аналогично полагается процедура для определения всех потоков x_{rs} , т. е.

$$x_{rs} = \min\{a_r^{(k)}, b_s^{(k)}\}, \quad (7.93)$$

где $a_r^{(k)} = a_r - \sum_{j=1}^{s-1} x_{rj}$; $b_s^{(k)} = b_s - \sum_{i=1}^{r-1} x_{is}$; k — номер итерации.

На следующей итерации выбирается переменная $x_{r, s+1}$, если предложение r -го источника реализовано не полностью, переменная $x_{r+1, s}$, если спрос s -го стока удовлетворен не полностью, и $x_{r+1, s+1}$ — в противном случае.

При построении начального опорного плана (выборе маршрута) методом минимального элемента используется информация о транспортных издержках $\bar{C} = \|c_{ij}\|_{m \times n}$.

В матрице издержек $\bar{C} = \|c_{ij}\|_{m \times n}$ нумеруются элементы в порядке их возрастания. Опорный план $\bar{X}$ заполняется перевозками в последовательности полученной нумерации по правилам, аналогичным методу «северо-западного угла».

Рассмотрим правила улучшения начального опорного плана.

Пусть имеется некоторый поток (план) и пусть $M(X)$ — множество дуг, для которых $x_{ij} > 0$:

$$M(X) = \{(i, j): X > 0\}.$$

Опорой потока $X = \{x_{ij}\}$ называется частичный граф $(V, M(X))$. Поток $X = \{x_{ij}\}$ невырожден, если его опора $(V, M(X))$ является связным графом (все элементы $x_{ij} > 0$). В случае вырожденности необходимо незначительно расширить исходную сеть, добавив маршрут с бесконечно малой перевозкой ε так, чтобы потоку $x(\varepsilon)$ в новой сети соответствовал уже связный частичный граф. Для этого в каждой компоненте связности выбирают произвольную вершину-источник и уменьшают ее интенсивность на бесконечно малую величину ε . После этого вводят фиктивную вершину с интенсивностью, равной суммарной интенсивности, на которую была

уменьшена интенсивность выбранных вершин. Фиктивную вершину соединяют с каждой из выбранных дугой с нулевым значением стоимости перевозки. Опора любого потока $x_{ij}(e)$ в новой сети будет связным графом.

Таким образом, для устранения вырождения достаточно соединить компоненты связности опоры нулевыми потоками.

Введем замену переменных: $v_i = -u_i$, $i = \overline{1, m}$. Тогда для оптимальности плана перевозок $\bar{X} = \|x_{ij}\|$ необходимо и достаточно существование потенциалов w_j , $j = \overline{1, n}$, пунктов потребления и потенциалов $(-u_i)$, $i = \overline{1, m}$, пунктов поставок таких, что

$$w_j - u_i \leq c_{ij}, \quad i = \overline{1, m}; \quad j = \overline{1, n}.$$

Для перевозок $x_{ij} > 0$ в силу теоремы о дополняющей нежесткости должны выполняться условия:

$$\text{для всех } x_{ij} > 0 \text{ обязательно } w_j - u_i = c_{ij}. \quad (7.94)$$

Разность $w_j - u_i$ можно рассматривать как приращение ценности единицы продукции при доставке ее из пункта отправления в пункт потребления. В связи с этим, если $w_j - u_i < c_{ij}$, то перевозка из пункта i в пункт j нерентабельна и должна отсутствовать, т. е. $x_{ij} = 0$. В противном случае перевозка рентабельна, т. е., если $w_j - u_i = c_{ij}$, то $x_{ij} > 0$.

Поскольку рентабельность зависит только от разности между потенциалами w_j и u_i , то, положив, например, $u_{i_0} = 0$ (i_0 — начальная вершина сети), можно вычислить потенциалы вершин сети по формулам $w_i = u_i + c_{ij}$, $u_i = w_j - c_{ij}$ в силу связности графа $(V, M(X))$.

Если для начального потока $X = \{x_{ij}\}$ будет выполнено условие

$$c_{ij} - (w_j - u_i) \geq 0, \quad (7.95)$$

то поток (план перевозок) является оптимальным. В противном случае план (поток) надо улучшить. Пусть для некоторой дуги $(r, l) \in M(X)$ условие (7.95) не выполняется. Из условия построения потенциалов каждой вершины следует, что существует цепь $[i_0, i_1, \dots, i_s = r]$, ведущая из вершины i_0 в вершину r , вдоль которой был определен потенциал u_s , и существует цепь $[i_0, i'_1, \dots, i'_k = l]$, ведущая из вершины i_0 в вершину l , вдоль которой был определен потенциал w_j . Частично эти цепи могут некоторыми вершинами совпадать.

Построим цикл. Для этого соединим указанные две цепи ребром $[r, l]$ и получим цикл

$$\xi = [i_0, i_1, \dots, i_s = r, i_{s+1} = l, i_{s+2}, \dots, i_t = i_0].$$

Направление дуги (r, l) должно определять направление обхода полученного цикла.

Пусть M_{ξ}^{+} и M_{ξ}^{-} — множества дуг, ориентированных в направлении обхода этого цикла и соответственно в противоположном.

Построим улучшенный поток $X^h = \{x_{ij}^h\}$ по правилу

$$x_{ij}^h = \begin{cases} x_{ij} + \Theta & \text{при } (i, j) \in M_{\xi}^+; \\ x_{ij} - \Theta & \text{при } (i, j) \in M_{\xi}^-; \\ x_{ij} & \text{при } (i, j) \in M_{\xi}^+ \cup M_{\xi}^-, \end{cases} \quad (7.96)$$

где

$$\Theta = \min \{ \min_{(i,j) \in M_{\xi}^-} x_{ij}, \min_{(i,j) \in M_{\xi}^+} (-x_{ij}) \}.$$

Разность между потоком, входящим в любую вершину, и потоком, выходящим из этой же вершины, при преобразовании (7.96) не изменится. Следовательно, поток (план) $X^h = \{x_{ij}^h\}$ будет выгоднее предыдущего потока $X = \{x_{ij}\}$ сети:

$$\begin{aligned} \sum_{(i,j) \in M} c_{ij} x_{ij} - \sum_{(i,j) \in M} c_{ij} x_{ij}^h &= \Theta \{ \sum_{(i,j) \in M_{\xi}^-} c_{ij} - \sum_{(i,j) \in M_{\xi}^+} c_{ij} \} = \\ &= \Theta (u_{i_0} - u_{i_1} + u_{i_1} - \dots - u_r - c_{rl} + w_l - w_{l+1} + w_{l+1} - \dots \\ &\quad \dots - w_{i_0}) = \Theta (w_l - u_r - c_{rl}) > 0. \end{aligned}$$

Если для полученного потока условие оптимальности не выполняется, то определяем новые потенциалы и улучшаем этот поток. Процедуру улучшения потока повторяем до тех пор, пока поток не станет оптимальным. Оптимальный поток соответствует оптимальному плану перевозок, т. е. решению задачи.

Рассмотрим транспортную задачу с ограниченными пропускными способностями.

Если на пропускную способность путей накладываются ограничения, то имеет место транспортная задача с ограниченными пропускными способностями. Такая задача описывается моделью (7.82) — (7.85), в которой ограничения (7.85) имеют вид

$$0 \leq x_{ij} \leq y_{ij}, \quad i = \overline{1, m}; \quad j = \overline{1, n}.$$

В этом случае, чтобы поток $X = \{x_{ij}\}$ был оптимальным, необходимо и достаточно существование для каждой вершины $i \in V$ такого потенциала u_i , что

$$\begin{aligned} w_j - u_i &\leq c_{ij}, \quad \text{если } x_{ij} = 0; \\ w_j - u_i &= c_{ij}, \quad \text{если } 0 < x_{ij} < y_{ij}; \\ w_j - u_i &\geq c_{ij}, \quad \text{если } x_{ij} = y_{ij}. \end{aligned} \quad (7.97)$$

После определения потенциалов по правилам, рассмотренным в предыдущем случае, при невыполнении условий (7.97) возможен один из трех случаев:

$$\begin{aligned} w_l - u_r &> c_{rl} \quad \text{при } x_{rl} = 0; \\ w_l - u_r &\neq c_{rl} \quad \text{при } 0 < x_{rl} < y_{rl}; \\ w_l - u_r &< c_{rl} \quad \text{при } x_{rl} = y_{rl}. \end{aligned}$$

Тогда улучшенный поток определяется по правилу (7.96), по величина Θ — по формуле

$$\Theta = \min \left\{ \min_{(i,j) \in M_{\xi}^{-}} x_{ij}, \min_{(i,j) \in M_{\xi}^{+}} (y_{ij} - x_{ij}) \right\},$$

при этом выполняются условия

$$0 \leq x_{ij}^h \leq y_{ij}, \quad (i, j) \in M.$$

Многие транспортные задачи включают пункты для перевозки грузов (пункты распределения, склады, оптовые базы и т. п.). В этом случае требуется отыскать план перевозок грузов, минимизирующий транспортные затраты при любой заданной схеме размещения перевалочных пунктов, т. е. рассматривается транспортная задача с промежуточными пунктами. Такую задачу также можно описать сетью, в которой вершины соответствуют источникам (пунктам поставки), стокам (пунктам потребления) и промежуточным пунктам (складам и т. п.).

Поскольку у источников имеется избыток груза, который необходимо вывезти, то соответствующая вершина характеризуется положительным числом, равным количеству вывозимого груза. Сток потребляет груз, и соответствующая вершина характеризуется отрицательным числом, равным количеству требуемого груза. Вершине, соответствующей промежуточному пункту, приписывается положительное число, обозначающее избыточное количество запаса груза или отрицательное число, обозначающее потребность в дополнительном грузе.

Задача с промежуточными пунктами легко сводится к классической транспортной задаче.

Для каждого промежуточного пункта k определяется чистый запас груза s_k . Если в k -м пункте имеется избыток запасов, то запас s_k должен быть положительным; если в k -м пункте требуется пополнение запасов, то s_k должен быть отрицательным числом.

Пусть a — фиктивный запас груза в каждом промежуточном пункте, т. е. запас, равный суммарному запасу, имеющемуся во всех пунктах. Для любого пункта k выполняются соотношения:

$$a_k = s_k + a; \quad b_k = a.$$

Для промежуточного пункта k вводится переменная потока x_{hk} при стоимости перевозки единицы груза $c_{hk} = 0$. При таком построении условие баланса не нарушается. Сумма всех a_i остается равной сумме всех b_j . Значение a_k должно быть настолько велико, чтобы оно могло включить любое количество груза, проходящего через промежуточный пункт в оптимальном режиме. Для этого запас a берется равным сумме запасов, имеющихся во всех пунктах. Поэтому разность $a - x_{hk}$ является общим количеством груза, перевезенного через промежуточный пункт k .

Пусть имеется сеть $G(V, L)$ с ограниченными пропускными способностями дуг, в которой один источник (вершина 0) и один сток (вершина $p+1$). Требуется определить максимальный поток из источника в сток для этой сети.

Математическая модель задачи имеет вид:

$$\text{максимизировать } f \quad (7.98) \\ \text{при ограничениях}$$

$$\sum_{(h,j) \in L} x_{hj} = f, \quad k = 0; \quad (7.99)$$

$$\sum_{(h,j) \in L} x_{hj} - \sum_{(i,j) \in L} x_{ih} = 0, \quad k = \overline{1, p}; \quad (7.100)$$

$$- \sum_{(i,p+1) \in L} x_{i,p+1} = -f, \quad k = p+1; \quad (7.101)$$

$$0 \leq x_{ij} \leq y_{ij}, \quad (i, j) \in L, \quad y_{ij} \geq 0, \quad (7.102)$$

где f — интенсивность источника (вершина 0); $-f$ — интенсивность стока (вершина $p+1$).

Максимальным является поток, для которого интенсивность f удовлетворяет соотношениям (7.98) — (7.102). Максимальное значение f называется *максимальным потоком* из источника 0 в сток $p+1$.

Пусть $S \subset V$ и $0 \in S, p+1 \notin S$. *Разрезом сети*, отделяющим источник 0 от стока $p+1$, называется множество дуг

$$L(S) = (S, \bar{S}) = \{(i, j) : i \in S, j \in \bar{S}\}.$$

Сумма пропускных способностей таких дуг $y(S, \bar{S})$ называется *величиной разреза*. Минимальным является разрез, у которого наименьшая пропускная способность.

Максимальный поток из источника 0 в сток $p+1$ равен пропускной способности минимального разреза, отделяющего источник 0 от стока $p+1$.

Если дуга (i, j) входит в минимальный разрез, то максимальный поток по дуге (i, j) равен пропускной способности y_{ij} этой дуги.

Вводится фиктивная дуга $(0, p+1)$, обладающая неограниченной пропускной способностью $y_{0,p+1} = \infty$.

Предполагается, что функция стоимости

$$c_{ij} = \begin{cases} 1, & i = 0, j = p+1; \\ 0, & i \neq 0, j \neq p+1. \end{cases} \quad (7.103)$$

Если поток $x_{0,p+1}$ минимальный, то через остальные дуги пройдет максимальное количество потока.

При $x_{0,p+1} > 0$ разность $w_{p+1} - u_0 = 1$ (где w_{p+1} и u_0 — потенциалы стока $p+1$ и источника 0 соответственно) по условиям теоремы двойственности. Потенциал одной из вершин может быть взят равным нулю, например $u_0 = 0$.

Двойственная задача к задаче минимизации потока:
максимизировать

$$f - \sum_{(i,j) \in G} z_{ij} y_{ij} \quad (7.104)$$

при ограничениях

$$w_j - u_i - z_{ij} \leq 0, \quad (i, j) \in G; \quad (7.105)$$

$$u_0 = 0, \quad w_{p+1} = 1. \quad (7.106)$$

Решение этой задачи следующее. Максимальный поток x^0_{ij} должен удовлетворять условиям

$$w_j^0 - u_i^0 = 0 \text{ при } 0 < x^0_{ij} < y_{ij}.$$

Поэтому все потенциалы и z_{ij} равны 0 или 1.

По теореме двойственности

$$x^0_{0,p+1} = f - \sum_{(i,j) \in L} z_{ij} y_{ij} = f - \sum_{(i,j) \in L_z^0} y_{ij}, \quad (7.107)$$

где $L_z^0 = \{(i, j): z_{ij} = 1\}$.

Из (7.107) следует, что максимальный поток равен $\sum_{L_z^0} y_{ij}$.

Любой путь от источника к стоку содержит дугу, входящую в разрез; L_z^0 является разрезом, так как если бы существовал путь от 0 к $p+1$, не содержащий элементов L_z^0 , то, увеличивая пропускные способности всех дуг, входящих в этот путь, можно было бы увеличить поток. Но $x^0_{ij} = 0$ при $(i, j) \in L_z^0$. Поэтому увеличение пропускной способности всех дуг, не входящих в L_z^0 , не изменит значений целевой функции двойственной задачи (7.104), а также максимального потока. Так как максимальный поток не больше пропускной способности любого разреза, то разрез L_z^0 является минимальным.

Приведем алгоритм определения максимального потока.

Шаг 1. Определить путь из источника 0 в сток $p+1$, по которому, поток принимает положительное значение. Если такого пути нет, перейти к шагу 3.

Шаг 2. Определить $\theta = \min \{y_{ij} > 0\}$, вычесть θ из всех y_{ij} для пути из 0 в $p+1$ и прибавить θ ко всем y_{ij} для пути из $p+1$ в 0. Перейти к шагу 1. Эта операция позволяет использовать оставшиеся пропускные способности и восстанавливает исходные пропускные способности, так как уменьшение пропускной способности в одном направлении увеличивает пропускную способность в противоположном направлении.

Шаг 3. Определить оптимальный поток $X = \|x_{ij}\|$, где

$$x_{ij} = \begin{cases} y_{ij} - y_{ij}^*, & y_{ij} > y_{ij}^*; \\ 0, & y_{ij} \leq y_{ij}^*; \end{cases} \quad (7.108)$$

максимальный поток

$$f = \sum_i x_{0i} = \sum_j x_{j, p+1}.$$

Здесь y_{ij} — исходные пропускные способности; y_{ij}^* — пропускные способности дуг, полученные на шагах 1 и 2.

Таким образом, максимальный поток f равен сумме всех положительных θ .

Пусть $y_{ij} = 1$ для всех $(i, j) \in L$.

Рассмотрим любой допустимый поток. Является этот поток максимальным или возможно его увеличение?

Шаг 1. Начиная с источника 0, отметить знаком «+» каждую дугу $(0, j)$ с нулевым потоком, а вершины j и 0 пометить знаком «*».

Шаг 2. Рассмотреть любой узел j , помеченный знаком «*». Поставить знак «+» на каждой дуге (j, k) с нулевым потоком, исходящей из вершины j , если вершина k не помечена, и пометить вершину k знаком «*». Поставить знак «-» на каждой дуге (k, j) с ненулевым потоком, входящей в вершину j , если вершина k не помечена, и пометить вершину k знаком «*». В конце шага 2 зачеркнуть знак «*» вершины j (это указывает на то, что вершина была просмотрена).

Шаг 3. Повторять операции шага 2, пока не будет помечена вершина $p+1$ или все помеченные вершины не будут просмотрены. Как только проставляется метка вершины $p+1$, происходит прорыв, так как обнаруживается путь, увеличивающий поток из вершины 0 в вершину $p+1$. Этот путь можно определить в явном виде с помощью знаков «+» или «-», проставленных на дугах, просматривая вершины в обратном направлении, начиная с вершины $p+1$. Добавить единичный поток по каждой дуге, помеченной знаком «+», уменьшить на единицу поток по каждой дуге, помеченной знаком «-». Вернуться к шагу 1. Если по окончании шага 2 вершина $p+1$ остается непомеченной, текущее решение определяет максимальный поток в сети.

Задача определения кратчайшего пути на сети

Задача определения кратчайшего пути заключается в нахождении самого экономического пути (по стоимости, продолжительности, длительности и т. п.) от начального пункта (источника) s_H до конечного (стока) s_K . В качестве начального пункта может быть задано целое множество источников, а в качестве конечного — множество стоков.

Пусть c_{ij} — длина каждой дуги и имеется путь из s_H в s_K

$$r = [i_0 = s_H, i_1, \dots, i_t = s_K] \quad (7.109)$$

длиной

$$l(r) = \sum_{\mu=1}^t c_{i_{\mu-1}i_{\mu}}. \quad (7.110)$$

Необходимо найти такой путь r , для которого $l(r) \rightarrow \min$.

Пусть s_H — источник единичной интенсивности, s_K — ее сток, остальные вершины — нейтральные; дуги имеют неограниченные пропускные способности, а стоимость перевозок равна длине соответствующей дуги.

Поток в такой сети должен удовлетворять соотношениям

$$\sum_i x_{ij} - \sum_j x_{ji} = \begin{cases} 1 & \text{при } i = s_H; \\ -1 & \text{при } i = s_K; \\ 0 & \text{при } i \neq s_H, i \neq s_K. \end{cases} \quad (7.111)$$

Необходимо найти такой поток, который удовлетворял бы (7.111) и минимизировал функционал

$$\sum_{(i,j) \in L} c_{ij} x_{ij}. \quad (7.112)$$

Таким образом, задача (7.109), (7.110) сводится к задаче (7.111), (7.112).

Пусть c_j — кратчайшее расстояние между вершиной s_H и вершиной j , причем $c_{s_H} = 0$. Необходимо найти c_{s_K} .

Кратчайшее расстояние между начальной вершиной и любой вершиной j находится как минимальная сумма длин путей, определяемых кратчайшим расстоянием до предыдущей вершины i , и расстояния между текущей вершиной j и предыдущей вершиной i , т. е.

$$c_j = \min_i \{c_i + c_{ij}\}. \quad (7.113)$$

Таким образом, кратчайшее расстояние до вершины j определяется после нахождения кратчайшего расстояния до каждой предыдущей вершины i , если i и j — смежные вершины, соединенные дугой.

Рассмотрим алгоритм определения кратчайшего пути.

Шаг 1. Для вершины $i_0 = s_n$ расстояние $c_{s_n} = c_{i_0} = 0$. Величина c_i включает расстояние до вершины i , которое используется для определения ближайшей вершины j .

Шаг 2. Пусть $i = i_0$. 2.1. Вычислить расстояние $c_j - c_i$ для всех вершин j . 2.2. Если $c_{ij} \geq c_j - c_i$ для всех j , то между вершинами i и j не существует более короткого пути. Если $i = s_n$, перейти к п.2.4; если нет, принять $i = i + 1$ и перейти к п.2.1. 2.3. Если $c_{ij} < c_j - c_i$, вычислить c_j' , используя формулу $c_j' = c_i + c_{ij}$, а затем определить c_j по формуле (7.113). Для $i = j$ заменить c_j и c_i на c_j' и c_i' . Если $i = s_n$, перейти к п.2.4, если нет, принять $i = i + 1$ и перейти к п.2.1. 2.4. Если значение c_j в п.2.3 изменилось, перейти к шагу 2, если нет, перейти к шагу 3.

Шаг 3. Полученные значения c_j определяют кратчайшее расстояние между вершинами i_0 и $j = \overline{1, s_n}$. Последняя дуга (i, j) в пути (i, j) должна удовлетворять условию

$$c_{i_1} = c_j - c_{i_1 j}.$$

После определения вершины i_1 предпоследняя вершина i_2 должна удовлетворять условию

$$c_{i_2} = c_{i_1} - c_{i_2 i_1}, \quad i_1 = j.$$

Процесс продолжается, пока не будет достигнута вершина $i_0 = s_n$.

Календарное планирование методом критического пути. Многие задачи осуществления некоторого производственного процесса требуют установить согласованную во времени последовательность выполнения различных его этапов. К таким задачам относятся разработка крупных проектов, например строительство дома, судна, автоматизированных систем управления, содержащих большое количество различных операций. Некоторые из этих операций могут выполняться одновременно, для других задается определенная последовательность. Такие операции, как внутренние отделочные работы и работы по благоустройству территории, можно совместить, другие, например возведение стен, можно начинать после закладки фундамента.

Задача управления состоит в том, чтобы обеспечить своевременное завершение проекта с учетом времени, необходимого для выполнения каждой операции, и определенных взаимосвязей, характеризующих последовательность их выполнения. При решении такой задачи необходимо определить операции (работы), являющиеся критическими для своевременного завершения всего проекта.

Каждый проект можно представить в виде графа, называемого *сетевым графиком*. Для его построения необходима следующая информация: множество всех операций, время выполнения каждой операции, последовательность их выполнения.

Каждой операции соответствует дуга ориентированного графа. Дуга (i, j) соединяет вершину i с вершиной j и в вершину i входят только дуги, соответствующие непосредственно предшествующим дугам данной операции. Часто в граф вводятся фиктивные дуги.

Вершины сетевого графика называются *событиями*.

Если все операции, которые отображаются дугами, входящими в соответствующую вершину, полностью завершены, говорят, что событие произошло.

Сетевой график не содержит контуров.

Пусть задан сетевой график $G=(V, L)$ с множеством событий V и множеством операций L , в котором содержится одно событие — источник (в него не входит ни одна дуга) и одно событие — сток (из него не выходит ни одна дуга). Такие события называются соответственно *начальным* и *конечным событием*.

Все события графа G нумеруются таким образом, чтобы для каждой дуги $(i, j) \in L$ соблюдалось условие $i < j$.

Пусть для события $i \in V$ имеем $T_H(i)$ и $T_K(i)$ — соответственно наиболее раннее и наиболее позднее время его возможного наступления, допускающее своевременное окончание всего проекта.

Наиболее ранние сроки наступления событий определяются следующим образом:

$$T_H(j) = \max_{i: (i, j) \in L} \{T_H(i) + t(i, j)\}, \quad (7.114)$$

где $t(i, j)$ — время выполнения работы (операции), соответствующей дуге $(i, j) \in L$.

Например (рис. 7.4),

$$T_H(9) = \max \{T_H(3) + t(3,9), T_H(4) + t(4,9), T_H(5) + t(5,9)\} = \\ = \max \{2 + 6, 6 + 3, 7 + 2\} = 9.$$

Наиболее поздние сроки наступления событий определяются следующим образом:

$$T_K(i) = \min_{j: (i, j) \in L} \{T_K(j) - t(i, j)\}. \quad (7.115)$$

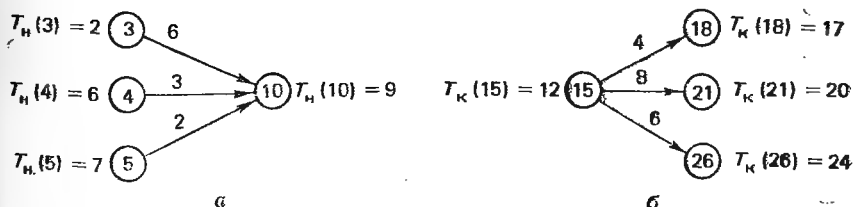


Рис. 7.4. Определение сроков наступления событий:

а — ранних; б — поздних

Например,

$$T_K(15) = \min\{T_K(18) - t(15, 18), T_K(21) - t(15, 21), \\ T_K(26) - t(15, 26)\} = \min\{17 - 4, 20 - 8, 24 - 6\} = 12.$$

Увеличение наиболее поздних сроков окончания всего проекта $T_K(n)$ на t единиц времени приведет к увеличению наиболее поздних сроков всех событий также на t единиц. Наиболее ранний срок $T_H(j)$ события j соответствует наибольшей длине пути от начального события (истока) к событию j , а разность $T_K(n) - T_K(j)$ — наибольшей длине пути от события j к конечному событию n .

Если $T_K(n) \geq T_H(n)$, то $T_K(j) \geq T_H(j)$ для всех событий j . Часто требуется определить максимально возможные длительности операций при условии своевременного окончания выполнения всего проекта. Поскольку любая операция (i, j) может начаться не ранее $T_H(i)$ и должна окончиться не позднее $T_K(j)$, то задержки выполнения всего проекта не будет, если на операцию $(i, j) \in L$ выделено $T_K(j) - T_H(i)$ единиц времени. Поэтому при выполнении операции (i, j) можно допустить максимальную задержку $T_K(j) - T_H(i) - t(i, j) \geq 0$. Величина $R_H(i, j) = T_K(j) - T_H(i) - t(i, j)$ называется *полным резервом времени операции* (i, j) .

Допустим, что на последующие операции (после операции (i, j)) не накладываются временные ограничения. Тогда на выполнение операции (i, j) можно выделить не более $T_H(j) - T_H(i)$ единиц времени.

Величина $R_C(i, j) = T_H(j) - T_H(i) - t(i, j)$ называется *свободным резервом времени операции* (i, j) . Он равен максимальной задержке выполнения операции (i, j) , не влияющей на выполнение последующих операций.

При отсутствии дополнительных временных ограничений на любую операцию проекта на выполнение операции (i, j) можно выделить не более $T_H(j) - T_K(i)$ единиц времени.

Величина $R_H(i, j) = T_H(j) - T_K(i) - t(i, j)$ называется *независимым резервом времени операции* (i, j) . При $R_H(i, j) < 0$ любая задержка в выполнении операции (i, j) приведет к дополнительным ограничениям на выполнение других операций.

Для любой операции (i, j)

$$R_H(i, j) \geq R_C(i, j) \geq R_H(i, j).$$

Операция называется *критической*, если любая задержка ее выполнения приводит к задержке выполнения всего проекта. Для критической операции $R_H(i, j) = 0$.

Задержки некритических операций, не превышающие их полный резерв времени, не могут нарушить своевременности окончания проекта. Следовательно, чтобы не нарушить срок завершения всего проекта, необходимо определить все критические операции проекта.

Путь, ведущий из начального события к конечному и состоящий только из критических операций, называется *критическим путем*.

Если $T_H(n) = T_K(n)$, то каждая операция пути наибольшей дли-

ны $T_n(n)$ от начального к конечному событию сетевого графика является критической.

Пример. Пусть задан сетевой график (рис. 7.5). Пути из начального события в конечное имеют следующие длины:

$$\begin{aligned}(1,2) - (2,5) &\rightarrow 3 + 6 = 9; \\(1,3) - (3,5) &\rightarrow 5 + 5 = 10; \\(1,4) - (4,5) &\rightarrow 5 + 2 = 7.\end{aligned}$$

Путь $(1,3) - (3,5)$ имеет наибольшую длину и является критическим.

При выполнении не критических операций путей $(1,2) - (2,5)$ и $(1,4) - (4,5)$ можно допустить некоторые задержки, не приводящие к превышению времени завершения проекта, равному 10 ед.

В случае неопределенности длительности выполнения операций для определения критического пути существует метод PERT (*Program Evaluation Review Technique*). Ожидаемое время выполнения операции (i, j) вычисляется на основании трех временных оценок по формуле

$$T_{ож} = \frac{A_{ij}}{6} + \frac{4}{6} B_{ij} + \frac{C_{ij}}{6},$$

где A_{ij} , B_{ij} , C_{ij} — соответственно оптимистическая, реалистическая и пессимистическая оценки времени выполнения. Дисперсия времени выполнения операции (i, j) определяется так:

$$D_{ij} = \left(\frac{C_{ij} - A_{ij}}{6} \right)^2$$

В этом случае $T_n(i)$ и $T_k(i)$ означают соответственно ожидаемый наиболее ранний и наиболее поздний сроки выполнения события i , $T_n(n)$ — ожидаемый наиболее ранний срок завершения всего проекта.

Обычно считают, что наиболее ранний срок завершения всего проекта характеризуется нормально распределенной случайной величиной с математическим ожиданием $T_n(n)$ и дисперсией, равной сумме дисперсий наиболее длинного пути от начального события i к конечному событию n .

Критический путь определяется следующим образом:

начиная с истока, для которого $T_n(0) = 0$, по формуле (7.114) определяются ранние сроки наступления всех событий j , включая событие-сток, т. е. $T_n(n)$;

начиная с последнего события-стока по формуле (7.115) определяются поздние сроки наступления всех событий i , включая событие-сток, т. е. $T_k(0)$;

операция (i, j) принадлежит критическому пути, если выполняются следующие условия:

$$\begin{aligned}T_n(i) &= T_k(i); \\T_n(j) &= T_k(j); \\T_n(j) - T_n(i) &= T_k(j) - T_k(i) = t(i, j),\end{aligned}$$

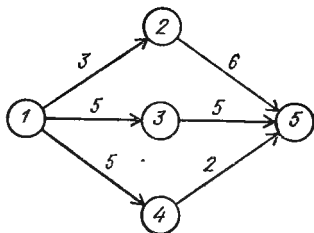


Рис. 7.5. Сетевой график выполнения проекта

т. е. для критических операций ранние сроки начала (окончания) и поздние сроки начала (окончания) совпадают (запас времени отсутствует). Поэтому критический путь определяет кратчайшую возможную продолжительность выполнения проекта.

7.3. ЦЕЛОЧИСЛЕННОЕ ПРОГРАММИРОВАНИЕ И КОМБИНАТОРНАЯ ОПТИМИЗАЦИЯ

Задачи комбинаторной оптимизации состоят в отыскании среди структуризованного конечного множества альтернатив наилучшего (относительно заданного критерия) решения.

Структурные особенности заданного конечного множества служат исходной предпосылкой для построения некоторого системного упорядоченного метода решения задачи, используемого вместо прямого перебора, и сравнения всех альтернативных вариантов.

С комбинаторной оптимизацией тесно связано целочисленное программирование, в задачах которого все или некоторые переменные должны принимать целочисленные значения. Требование целочисленности решения порождает нелинейный характер задачи. Но во многих случаях, если нелинейность связана только с условиями целочисленности результата решения, целочисленное программирование рассматривается как раздел математической оптимизации линейных моделей, в которых на некоторые или все переменные накладываются условия целочисленности.

Целочисленная и комбинаторная оптимизации имеют необычайно широкую область применения. Часто задачи, связанные с принятием плановых решений, требуют целочисленности вектора управляемых переменных. К таким задачам относятся задачи составления последовательности производственных процессов, календарного планирования работы организаций, планирования и обеспечения материально-технического снабжения, размещения предприятий, распределения ресурсов, планирования использования оборудования, выбора маршрута, упорядочения и др. К комбинаторной оптимизации кроме задач упорядочения, календарного планирования и выбора маршрута относятся задачи конструирования вычислительных машин, кодирования сигналов и др.

Задачи целочисленного линейного программирования описываются моделью следующего вида:
оптимизировать

$$\sum_{j=1}^n c_j x_j \quad (7.116)$$

при условиях

$$\sum_{j=1}^n a_{ij} x_j \leq a_{i0}, \quad i = \overline{1, m}; \quad (7.117)$$

$$x_j \geq 0, \quad j = \overline{1, n}; \quad (7.118)$$

$$x_j — \text{целые}, \quad j = \overline{1, k} \quad (k \leq n). \quad (7.119)$$

Если $k=n$, модель (7.116)—(7.119) описывает полностью целочисленную задачу линейного программирования, если $k < n$, то — частично целочисленную.

В зависимости от содержания задачи оптимизация целевой функции (7.116) может соответствовать максимизации или минимизации.

Модель задачи целочисленного программирования (7.116)—(7.119) отличается от модели задачи линейного программирования только наличием дополнительного ограничения (7.119). Это ограничение приводит к тому, что максимальное значение целевой функции задачи целочисленного программирования меньше значения целевой функции соответствующей задачи линейного программирования.

Одна из важных отличительных черт между задачами целочисленного и линейного программирования — отсутствие простого способа, позволяющего определять оптимальность допустимого решения.

В задачах целочисленного программирования вычисления гораздо сложнее, чем в задачах линейного программирования. Поэтому при использовании ЭВМ для решения задач целочисленного программирования большой размерности возникают большие сложности, связанные с ошибками округления.

Часто исследователи прибегают к приближенному результату решения, полученного путем округления до целых значений результатов решения задачи линейного программирования (7.116)—(7.118). При округлении обычно нарушается допустимость решения, т. е. не выполняются ограничения вида (7.117).

Все методы решения задач целочисленного программирования делятся на две группы: методы отсечения и комбинаторные методы. Эти методы исключают необходимость явного перебора всех допустимых альтернатив, обеспечивают частичный перебор сравнительно небольшого числа допустимых вариантов и неявный перебор всех остальных.

Большинство методов обеих групп базируется на процедурах формирования и решения последовательности взаимосвязанных задач, в основе которых лежит исходная задача, т. е. на каждой последующей итерации рассматривается задача (методы отсечения) или задачи (комбинаторные методы), которые порождаются задачей (задачами) предыдущей итерации. Порожденным задачам соответствуют ослабленные задачи, получаемые отбрасыванием условий целочисленности решения, что приводит к расширению области допустимых решений. Порожденные задачи должны обладать тем свойством, что оптимальное решение хотя бы одной из них должно совпадать с оптимальным решением задачи, породившей ее.

В методах отсечения на каждой итерации добавляется линейное ограничение, удовлетворяющее целочисленному решению исходной задачи и исключающее текущее нецелочисленное решение предыдущей итерации. Вычислительный процесс заканчивается при достижении целочисленного решения. Сходимость достигается

ся за конечное число итераций, но это число может быть очень большим.

Все комбинаторные методы базируются на идее перебора всех допустимых целочисленных решений.

В методе ветвей и границ, относящемся к комбинаторным методам, на каждой итерации строятся две задачи, в основе которых лежит ослабленная задача предыдущей итерации. Это построение соответствует делению пространства решений исходной задачи и отбрасыванию области, не содержащей допустимого целочисленного решения.

Метод неявного перебора вариантов, предназначенный для решения задач с булевыми переменными, основывается на идеях метода ветвей и границ.

Часто пользуются другими подходами к решению целочисленных задач, позволяющими находить решения, близкие к оптимальному. К таким подходам относится использование случайной выборки допустимых решений с последующим улучшением каждого попавшего в выборку решения, если способ улучшения просто определяется.

Существует подход к решению задач, названный *эвристическим программированием*, который заключается в применении некоторых правил, основанных на рациональных соображениях, методе проб и ошибок для отыскания допустимого, близкого к оптимальному, решения. Эти правила применяются сообразно с конкретной ситуацией и не могут быть приведены в стройную систему.

Метод отсекающих плоскостей. Пусть задана полностью целочисленная задача, которая описывается моделью:
максимизировать

$$\sum_{j=1}^n c_j x_j \quad (7.120)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j = a_{i0}, \quad i = \overline{1, m}; \quad (7.121)$$

$$x_j \geq 0, \quad j = \overline{1, n}; \quad (7.122)$$

$$\text{все } x_j \text{ — целые.} \quad (7.123)$$

Для полностью целочисленных задач математик Гомори разработал следующий метод. Вначале рассматривается выпуклое множество допустимых решений, определенное линейными ограничениями и условиями неотрицательности переменных исходной задачи (7.121) — (7.122), и находится экстремальная точка этого множества. Если такое решение оказывается нецелочисленным, то добавляют ограничение, отсекающее текущую экстремальную точку и уменьшающее выпуклое множество, но не отсекающее ни одной экстремальной точки выпуклой оболочки множества допустимых целочисленных решений. Процедура повторяется, т. е. строится такое число дополнительных ограничений, пока экстремальная точка усеченного выпуклого множества допустимых решений

задачи линейного программирования не совпадает с экстремальной точкой выпуклой оболочки.

Это дополнительное линейное ограничение, которому должно удовлетворять любое решение, допустимое по условиям (7.121) — (7.123), строится следующим образом.

Просуммируем по всем i правые и левые части условий (7.121):

$$\sum_{j=1}^n a_{ij} x_j = a_{i0}, \quad (7.124)$$

где

$$a_j = \sum_{i=1}^m a_{ij}, \quad a_0 = \sum_{i=1}^m a_{i0}.$$

Из соотношения (7.124) и условий (7.122), (7.123) следует

$$\sum_{j=1}^n [a_j] x_j \leq a_0, \quad (7.125)$$

где $[a_j]$ — целая часть числа a_j , т. е. наибольшее число, меньшее или равное действительному числу a_j .

Поскольку выражение в правой части (7.125) — целое число, то

$$\sum_{j=1}^n [a_j] x_j \leq [a_0] \quad (7.126)$$

или

$$\sum_{j=1}^n [a_j] x_j + y = [a_0], \quad (7.127)$$

где $y \geq 0$ и целое число.

Если к ограничениям (7.121) добавить ограничение (7.127), названное *отсечением в целых частях*, исходная задача останется целочисленной.

При решении задач удобнее пользоваться отсечением в дробных частях, которое легко получается из (7.127):

$$\sum_{j=1}^n (-d_j) x_j + x = -d_0, \quad (7.128)$$

где d_j и d_0 удовлетворяют соотношениям:

$$\begin{aligned} [a_j] + d_j &\equiv a_j \text{ и } [a_0] + d_0 \equiv a_0; \\ 0 &\leq d_j < 1; \quad 0 \leq d_0 < 1; \\ j &= \overline{1, n}. \end{aligned}$$

Алгоритм отсекающих плоскостей содержит три шага.

Шаг 1. Решить задачу линейного программирования (7.120) — (7.122), условия (7.123) отбрасываются.

Шаг 2. Если полученное решение является целочисленным, прекратить вычисления, так как оно является решением исходной задачи (7.120) — (7.123). Если же нет, выбрать одну из дробных базисных переменных и составить отсечение в дробных частях (7.128) из ограничения, содержащего эту базисную переменную в текущем оптимальном решении задачи линейного программирования.

Шаг 3. Расширить исходную задачу линейного программирования полученным новым ограничением (отсечением). Решить расширенную задачу и перейти к шагу 2.

Пусть на последней итерации x_s — базисная переменная в i -й строке симплекс-таблицы:

$$x_s + \sum_{j \in J} p_{ij} x_j = a_0, \quad (7.129)$$

где J — множество индексов небазисных переменных; a_0 — имеет дробное значение; p_{ij} — коэффициенты при небазисных переменных в i -й строке.

На шаге 2 отсечение в дробных частях:

$$\sum_{j \in J} (-d_{ij}) x_j + y = -d_0, \quad (7.130)$$

где d_{ij} — дробная часть p_{ij} , а d_0 — дробная часть a_0 .

Текущее решение задачи линейного программирования не удовлетворяет новому ограничению (7.130), так как $y = -d_0$ (отрицательно). При расширении задачи соотношением (7.130) и требовании $y \geq 0$ текущее решение задачи линейного программирования становится недопустимым, т. е. дробное решение отсекается и область допустимых решений сужается. Отсюда следует, что на шаге 3 значение целевой функции может только уменьшиться, а не возрасти.

Для решения расширенной задачи линейного программирования на каждой итерации удобно пользоваться двойственным симплексным методом.

Если в решении имеется несколько дробных управляемых переменных, то для построения отсечения выбирается переменная с наибольшей дробной частью, так как при таком выборе переменной сходимость к решению часто повышается. Обычно число ограничений-отсечений, за счет которых расширяется исходная задача, не превосходит $m+n$.

Метод ветвей и границ. Пусть требуется решить задачу (7.120) — (7.123), к которой добавлены ограничения вида

$$M_j \leq x_j \leq N_j, \quad j = \overline{1, n}. \quad (7.131)$$

Метод ветвей и границ базируется на следующей идее.

Введем целое число K , удовлетворяющее соотношению

$$M_j \leq K \leq N_j - 1.$$

Оптимальное решение задачи (7.120) — (7.123), (7.131) должно удовлетворять одному из двух линейных ограничений:

$$1) \quad x_j \geq K + 1; \quad (7.132)$$

$$2) \quad x_j \leq K. \quad (7.133)$$

Допустим, решение задачи линейного программирования (7.120) — (7.122), (7.131) содержит $x_1 = 1\frac{1}{2}$.

Решим две задачи линейного программирования (7.120) — (7.122), одна из которых дополнена условием $2 \leq x_1 \leq N_1$, а другая — условием $M_1 \leq x_1 \leq 1$. Пусть решение каждой из этих задач целочисленно, т. е. выполняются условия (7.123). Оптимальным решением исходной задачи целочисленного программирования является то решение, при котором значение целевой функции (7.120) будет большим.

Приведем алгоритм метода ветвей и границ.

На первой итерации алгоритма решаем исходную задачу линейного программирования (7.120) — (7.122), (7.131). Значение целевой функции (7.120)

принимая в качестве верхней оценки ξ_0^1 целевой функции последующих итераций.

На второй итерации формируем две задачи, каждая из которых содержит исходную задачу, дополненную одним из ограничений — (7.132) или (7.133), т. е. в списке задач имеется две задачи, которые необходимо решить и сравнить соответствующие значения целевой функции каждой задачи с имеющейся ее оценкой.

Пусть выполнено $t-1$ итераций. Итерация t содержит следующие три шага.

Шаг 1. Проверить, имеются ли в списке задачи. Если задач нет, прекратить вычисления, зафиксировать целочисленное решение (если оно имеется) с наилучшей оценкой ξ_0^t . В противном случае выбрать одну задачу из списка.

Шаг 2. Решить выбранную из списка задачу. Если задача не имеет допустимого решения или целевая функция оптимального решения меньше или равна нижней оценке ξ_0^{t-1} , то эта задача считается прозондированной, ее исключить из рассмотрения, принять $\xi_0^{t+1} = \xi_0^t$ и перейти к шагу 1. Если оптимальное решение задачи линейного программирования удовлетворяет условию целочисленности, зафиксировать его, принять ξ_0^{t+1} равным оптимальному значению функции и вернуться к шагу 1. В противном случае перейти к шагу 3.

Шаг 3. В векторе управления $\bar{X} = (x_1, \dots, x_n)$ выбрать любую нецелочисленную переменную $x_j = d_j$, $j = \overline{1, n}$, сформировать две новые задачи линейного программирования, добавив к задаче, выбранной на шаге 1, ограничение, в котором нижняя граница x_j заменена на $[d_j] + 1$, а к другой — ограничение, в котором верхняя граница x_j заменена на $[d_j]$, т. е. в одной задаче $x_j \geq [d_j] + 1$, в другой $x_j \leq [d_j]$.

Принять $\xi_0^{t+1} = \xi_0^t$ и вернуться к шагу 1.

Таким образом, процесс ветвления продолжается до тех пор, пока каждая из задач будет иметь целочисленное решение или улучшить полученное решение невозможно.

Метод неявного перебора вариантов. Метод неявного частично-перебора вариантов применим к решению задач, модель которых имеет вид:

максимизировать

$$\sum_{j=1}^n c_j x_j \quad (7.134)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j \leq a_{i0}, \quad i = \overline{1, m}; \quad (7.135)$$

$$x_j = \begin{cases} 0, & j = \overline{1, n}. \end{cases} \quad (7.136)$$

Предполагается, что каждый коэффициент c_j — рациональное целое число.

Полностью целочисленные задачи всегда можно преобразовать к задаче с булевыми переменными, так как если для любой целочисленной переменной $0 \leq z \leq k$, а $z_1, z_2, \dots, z_k$ — булевы переменные, т. е.

$$z_s = \begin{cases} 0, & s = \overline{1, k}, \\ 1, & \end{cases}$$

то

$$z = z_1 + z_2 + \dots + z_k$$

или

$$z = 1z_1 + 2z_2 + 2^2z_3 + \dots + 2^{p-1}z_p,$$

где p — наименьшее целое число, определяемое условием

$$2^p - 1 \geq k.$$

Алгоритм, реализующий метод частичного перебора вариантов, базируется на операциях сложения (вычитания). Поэтому такой метод часто называют *аддитивным алгоритмом*.

Ограничения булевости переменных (7.136) определяют 2^n возможных выборов значений вектора управляемых переменных $\bar{X} = (x_1, \dots, x_n)$. Ограничения (7.135) делают многие из этих вариантов недопустимыми; только некоторые из допустимых вариантов являются оптимальными.

Подмножество переменных x_j , в котором каждому x_j соответствует числовое значение, равное нулю или единице, называется *частичным решением*. Те переменные x_l , $l \neq j$, которые не входят в частичное решение, называются *свободными переменными*. Любой набор свободных переменных x_l , $l \neq j$, каждая из которых принимает конкретное значение, равное нулю или единице, называется *дополнением соответствующего частичного решения*. Если в частичное решение входит k переменных, то существует 2^{n-k} дополнений.

В аддитивном алгоритме возможные дополнения порождают ветви дерева, частичное решение — вершины.

Пусть $n=5$ и частичное решение ($x_2=0$, $x_4=1$). Свободными переменными являются x_1 , x_3 , x_5 . Существует $2^{5-2}=8$ дополнений частичного решения, содержащего две переменные; одним из дополнений, например, является $x_1=0$, $x_3=0$, $x_5=1$.

Предположим, что известна нижняя оценка ξ^0 оптимального значения целевой функции и соответствующее допустимое решение.

Допустим построено некоторое частичное решение.

Если можно показать, что не существует допустимого дополнения к этому частичному решению, при котором значение целевой функции существует и превышает нижнюю оценку ξ^0 , то нет необходимости дальнейшего ветвления, т. е. зафиксированное частичное решение не может содержаться в оптимальном решении задачи и нет смысла рассматривать все его возможные дополнения. Такое частичное решение отбрасывается как неперспективное и является прозондированным.

Таким образом, при зондировании частичного решения, содержащего k переменных, неявным образом перебирается 2^{n-k} возможных вариантов решений, удовлетворяющих условию булевости (7.136).

Частичные решения зондируются по следующему правилу.

Частичное решение является прозондированным, если не выполняется соотношение

$$\sum_{j \in S} a_{ij}x_j \leq a_{i0} - \sum_{j \in L} a_{ij}x_j, \quad i = 0, 1, 2, \dots, m, \quad (7.137)$$

т. е. не существует допустимого дополнения, которому соответствует значение целевой функции, превосходящее нижнюю оценку при выполнении соотношения

$$\sum_{j \in S} \min(a_{ij}, 0) > a_{i0} - \sum_{j \in L} a_{ij}x_j, \quad i = 0, 1, 2, \dots, m, \quad (7.138)$$

где $a_{0j} = -c_j$ и $a_{00} = -\xi^t - 1$; S — подмножество индексов свободных переменных; L — подмножество индексов переменных частичного решения, которое зондируется.

Если установлен факт существования допустимого дополнения, улучшающего значение целевой функции, то иногда можно просто определить значения переменных этого дополнения по следующему правилу.

Если для свободной переменной x_h выполняется соотношение

$$\sum_{j \in S} \min(a_{ij}, 0) + |a_{ih}| > a_{i0} - \sum_{j \in L} a_{ij}x_j \quad \text{при любом } i, \quad (7.139)$$

то

$$x_h = \begin{cases} 0, & \text{если } a_{ih} > 0; \\ 1, & \text{если } a_{ih} < 0. \end{cases} \quad (7.140)$$

Иногда более удобно зондирование частичного решения по правилу (7.138) и проверку выполнения соотношения (7.140) проводить для составных ограничений, образованных по правилу

$$\sum_{j=1}^n \left(\sum_{i=1}^m y_i a_{ij} \right) x_j \leq \sum_{i=1}^m y_i a_{i0}, \quad (7.141)$$

где $y_j > 0$. Такой подход позволяет сократить количество проверяемых ограничений (7.135), т. е. ведет к уменьшению затрат машинного времени.

С помощью рассмотренного метода можно решать задачи, содержащие не более 100 переменных и не более 50 ограничений.

Рассмотрим алгоритм реализации метода неявного перебора вариантов. Перед первой итерацией в список задач помещают две задачи, соответствующие двум частичным решениям: $x_q = 0$ и $x_q = 1$.

При задаче максимизации индекс q определяют по правилу

$$q = \{j | \max c_j, \quad j = \overline{1, n}\}. \quad (7.142)$$

Шаг 1. Проверить, имеются ли задачи в списке. Если имеются, выбрать одну из них для зондирования, вычеркнуть ее из списка. В противном случае прекратить вычисления.

Шаг 2. Прозондировать выбранное частичное решение по правилу (7.138). Если не существует допустимого дополнения, для которого значение целевой функции превосходит достигнутый рекорд ξ^t (наилучшее значение целевой функции, достигнутое на предыдущей итерации), то частичное решение прозондировано, полагают $\xi^{t+1} = \xi^t$ и возвращаются к шагу 1.

Если частичное решение не прозондировано, его необходимо расширить за счет свободных переменных по правилу (7.139) — (7.140) и перейти к шагу 3.

Шаг 3. Если частичное решение (расширенное) содержит все n переменных, зафиксировать его, определить соответствующее значение целевой функции (рекорд) ξ^{t+1} и вернуться к шагу 1. В противном случае перейти к шагу 4.

Шаг 4. Расширить частичное решение за счет некоторой свободной переменной x_q , образуя две задачи, первая из которых соответствует частичному решению, расширенному за счет переменной $x_q = 0$ ($x_q = 1$), вторая — за счет переменной $x_q = 1$ ($x_q = 0$). Принять $\xi^{t+1} = \xi^t$ и вернуться к шагу 1.

Таблица 7.14

C			1	1	0	0
	B_X	A_0	A_1	A_2	A_3	A_4
0	x_3	75	20	10	1	0
0	x_4	55	12	7	0	1
	Δ	0	-1	-1	0	0

После прекращения вычислений на шаге 1 в случае, когда зафиксировано допустимое решение и рекорд ξ^t , это решение является оптимальным. В противном случае допустимого решения не существует.

Пример. Дана задача, описываемая следующей моделью:
максимизировать

$$(x_1 + x_2)$$

при ограничениях

$$20x_1 + 10x_2 \leq 75; \quad 12x_1 + 7x_2 \leq 55; \quad x_1 \geq 0, \quad x_2 \geq 0; \quad x_1, x_2 — \text{целые.}$$

Решение. Перепишем модель задачи в канонической форме:
максимизировать

$$(x_1 + x_2)$$

при ограничениях

$$20x_1 + 10x_2 + x_3 = 75; \quad 12x_1 + 7x_2 + x_4 = 55;$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0;$$

$$x_1, x_2, x_3, x_4 — \text{целые.}$$

Отбросив условия целочисленности, решим задачу линейного программирования методом симплекс-таблиц. Табл. 7.14 — исходная симплекс-таблица.

Проведя симплексные преобразования, получим табл. 7.15.

Поскольку все элементы индексной строки положительные, то получено оптимальное решение задачи линейного программирования: $\bar{X} = \left(0, \frac{15}{2}, 0, \frac{5}{2}\right)$

и значение целевой функции $a_{00} = \frac{15}{2}$.

Решение нецелочисленно. Для нахождения целочисленного решения можно использовать метод отсекающих плоскостей либо метод ветвей и границ.

Метод отсекающих плоскостей

Построим правильное отсечение по переменной $x_2 = \frac{15}{2}$:

$$d_{20} = x_{20} - [x_{20}] = \frac{15}{2} - 7 = \frac{1}{2}; \quad d_{23} = x_{23} - [x_{23}] = \frac{1}{10} - 0 = \frac{1}{10}.$$

Таблица 7.15

C			1	1	0	0
	B_X	A_0	A_1	A_2	A_3	A_4
1	x_2	$\frac{15}{2}$	2	1	$\frac{1}{10}$	0
0	x_4	$\frac{5}{2}$	-2	0	$-\frac{7}{10}$	1
	Δ	$\frac{15}{2}$	1	0	$\frac{1}{10}$	0

Таблица 7.16

C			1	1	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
1	x_2	$\frac{15}{2}$	2	1	$\frac{1}{10}$	0	0
0	x_4	$\frac{5}{2}$	-2	0	$-\frac{7}{10}$	1	0
0	x_5	$-\frac{1}{2}$	0	0	$-\frac{1}{10}$	0	1
	Δ	$\frac{15}{2}$	1	0	$\frac{1}{10}$	0	0
	Θ				1		

Тогда отсечение имеет вид: $-\frac{1}{2} = x_5 - \frac{1}{10}x_{53}$. В результате получена табл. 7.16.

Находим решение с помощью двойственного симплекс-метода. Направляющий элемент x_{53} , так как только $x_5 < 0$ и $\Theta_{\min} = \Theta_3 = 1$.

Проделав симплексные преобразования, получим табл. 7.17.

Получено оптимальное решение, так как все элементы индексной строки неотрицательные. Решение $\bar{X}_0 = (0, 7, 5, 6)$ — целочисленно. Поэтому оптимальным решением исходной задачи является: $x_1 = 0$; $x_2 = 7$. Значение целевой функции $f_{\max} = 7$.

Метод ветвей и границ

Верхняя оценка на множестве решений G^0 равна $\xi(G^0) = \frac{15}{2}$. Разбиваем исходное множество G^0 на два подмножества:

$$G_1^I = \left\{ \bar{X} : \bar{X} \in G^0 \cap x_2 \leq \left\lfloor \frac{15}{2} \right\rfloor \right\};$$

$$G_2^I = \left\{ \bar{X} : \bar{X} \in G^0 \cap x_2 \geq \left\lfloor \frac{15}{2} \right\rfloor + 1 \right\}.$$

Находим решения и оценки на этих подмножествах.

Строим две задачи линейного программирования добавлением к исходной задаче ограничений $x_2 \leq 7$ и $x_2 \geq 8$, т. е. $x_2 + x_5 = 7$ и $-x_2 + x_5' = -8$. Полученные ограничения дописываем к табл. 7.15 и получим табл. 7.18.

Таблица 7.17

C			1	1	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
1	x_2	7	2	1	0	0	1
0	x_4	6	-2	0	0	1	-7
0	x_3	5	0	0	1	0	-10
	Δ	7	1	0	0	0	1

Таблица 7.18

C			1	1	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
1	x_2	$\frac{15}{2}$	2	1	$\frac{1}{10}$	0	0
0	x_4	$\frac{5}{2}$	-2	0	$-\frac{7}{10}$	1	0
	Δ	$\frac{15}{2}$	1	0	$\frac{1}{10}$	0	0
0	x_5	7	0	1	0	0	1
0	$\tilde{x}_5$	$-\frac{1}{2}$	-2	0	$-\frac{1}{10}$	0	1
	Θ		$\frac{1}{2}$		1		
0	x'_5	-8	0	-1	0	0	1
0	$\tilde{x}'_5$	$-\frac{1}{2}$	2	0	$\frac{1}{10}$	0	1
	Θ'						

Вместо строки $x_5(x'_5)$ записываем строку $\tilde{x}_5(\tilde{x}'_5)$, представляющую собой линейную комбинацию строк x_2 и $x_5(x'_5)$.

Решаем полученные задачи двойственным симплекс-методом.

Поскольку в строке x'_5 имеем $x_{50} < 0$, а все $x_{5j} \geq 0$, то на множестве G_2^I задача неразрешима.

Прделав симплексные преобразования с направляющим элементом $x_{31} = -2$, получим табл. 7.19.

Полученное решение нецелочисленно.

Строим две задачи, дополняя предыдущую ограничением $x_1 \leq \left[\frac{1}{4} \right]$ или $x_1 \geq \left[\frac{1}{4} \right] + 1$, т. е. разбиваем множество G_1^I на два подмножества:

Таблица 7.19

C			1	1	0	0	0
	B_X	A_0	A_1	A_2	A_3	A_4	A_5
1	x_2	7	0	1	0	0	1
0	x_4	3	0	0	$-\frac{3}{5}$	1	-1
1	x_1	$\frac{1}{4}$	1	0	$\frac{1}{20}$	0	$-\frac{1}{2}$
	Δ	$\frac{29}{4}$	0	0	$\frac{1}{20}$	0	$\frac{1}{2}$

$$G_1^2 = \left\{ \bar{X} : \bar{X} \in G_1^1 \cap x_1 \leq \left[\frac{1}{4} \right] \right\};$$

$$G_2^2 = \left\{ \bar{X} : \bar{X} \in G_1^1 \cap x_1 \geq \left[\frac{1}{4} \right] + 1 \right\}.$$

На множестве G_1^2 получается решение $x_1 = 0$; $x_2 = 7$ и значение целевой функции равно 7, т. е. $\xi(G_1^2) = 7$. Дальнейшие вычисления прекращаются, так как не может быть решения с лучшей оценкой, ибо $\xi(G^0) = \frac{15}{2}$. Таким образом, решение исходной задачи следующее:

$$x_1 = 0; x_2 = 7 \text{ и } f_{\max} = 7.$$

7.4. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Строго линейные модели, в которых использовалась пропорциональность, линейность и аддитивность, являются далеко не адекватными многим реальным ситуациям. В действительности такие зависимости, как общие затраты, выпуск продукции и т. п., от плана производства $\bar{X} = (x_1, \dots, x_n)$ носят нелинейный характер.

Часто применение моделей линейного программирования в условиях нелинейности является успешным. Поэтому необходимо определять, в каких случаях линейаризованный вариант задачи является адекватным отображением нелинейного явления.

Если имеются сведения относительно допустимого диапазона значений переменных в оптимальном решении, то, как правило, можно построить соответствующие ограничения и получить достаточно надежное линейное приближение. В тех же случаях, когда существует широкий диапазон допустимых решений и нет сведений о характере оптимального решения, построить достаточно хорошее линейное приближение нельзя.

Значимость нелинейного программирования и его использование постоянно возрастают.

Часто нелинейности в моделях обуславливаются эмпирическими наблюдениями соотношений, таких как непропорциональные изменения затрат, выхода продукции, показателей качества или структурно полученные соотношения, к которым относятся постулируемые физические явления, а также выведенные математически или установленные руководством правила поведения.

Множество разнообразных обстоятельств приводит к нелинейной формулировке ограничений или целевых функций. При небольшом количестве нелинейностей или, если нелинейности не существенны, увеличение объема вычислений может быть незначительным.

Всегда необходимо проанализировать размерность и сложность модели и оценить влияние линейаризации на принимаемое решение.

Часто пользуются двухэтапным подходом к решению задач: строят нелинейную модель небольшой размерности, находят область, содержащую ее оптимальное решение, а затем используют более детальную модель линейного программирования большей

размерности, аппроксимация параметров которой базируется на полученном решении нелинейной модели.

Для решения задач, описываемых нелинейными моделями, не существует такого универсального метода решения, как симплекс-метод решения задач линейного программирования.

Какой-либо метод нелинейного программирования может оказаться весьма неэффективным для решения задач одного типа и совершенно неприемлемым для решения других задач.

Большинство методов нелинейного программирования не всегда обеспечивает сходимость за конечное число итераций. Некоторые методы обеспечивают монотонное улучшение значения целевой функции при переходе от одной итерации к другой.

Общая задача нелинейного программирования описывается моделью следующего вида:

минимизировать (максимизировать) целевую функцию

$$f_0(x_1, x_2, \dots, x_n) \quad (7.143)$$

при ограничениях

$$f_i(x_1, x_2, \dots, x_n) \leq 0, \quad i = \overline{1, m}, \quad (7.144)$$

где $f_0(x_1, \dots, x_n)$ и $f_i(x_1, \dots, x_n)$, $i = \overline{1, m}$, — действительные нелинейные, но регулярные функции n действительных переменных.

Итак, общая задача нелинейного программирования формулируется следующим образом: требуется найти n -мерный вектор $\bar{X} = (x_1, \dots, x_n)$, который минимизирует (максимизирует) целевую функцию (7.143) и удовлетворяет ограничениям (7.144).

В общем случае область допустимых решений, описываемая ограничениями (7.144), может быть не только невыпуклой, но и несвязной; линия уровня функции цели также может состоять из отдельных кусков, т. е. геометрическая интерпретация задачи (7.143)—(7.144) следующая: функция цели описывает своеобразную некоторую местность с оврагами, возвышенностями и т. п., а ограничения описывают некоторые изгороди на этой местности; необходимо отыскать самую низкую (высокую) точку местности на территории, окруженной изгородями.

Наименьшее значение линии уровня функции цели, после которого общих точек линии уровня и области допустимых решений не существует, является оптимальным значением функции цели, а общие точки линии уровня и области допустимых решений являются решением задачи нелинейного программирования.

Решение задачи может быть внутри области допустимых решений или на ее границе.

Функция цели обычно имеет несколько локальных и глобальных экстремумов, что затрудняет решение задачи.

Точка области допустимых решений является глобальным (абсолютным) минимумом целевой функции, если ни в одной другой точке этой области целевая функция не принимает меньшего значения.

Точка области допустимых решений является локальным (относительным) минимумом, если ни в какой точке малой ее области

целевая функция не принимает меньшего значения, т. е. $\bar{X}_0$ — точка локального минимума целевой функции $f_0(\bar{X})$, если $f_0(\bar{X}_0) \leq f_0(\bar{X})$ для всех точек $\bar{X}$ таких, что $|\bar{X} - \bar{X}_0| \leq \varepsilon$ (где $\varepsilon > 0$), — произвольно малое число.

Глобальный минимум совпадает с одним из локальных, но локальный минимум не всегда совпадает с глобальным.

В задачах выпуклого программирования локальный и глобальный экстремумы совпадают.

Функция $f(\bar{X})$ называется *выпуклой вниз* (выпуклой), если

$$f(k\bar{X}_1 + (1-k)\bar{X}_2) \leq kf(\bar{X}_1) + (1-k)f(\bar{X}_2)$$

для любых точек $\bar{X}_1, \bar{X}_2$ и $0 \leq k \leq 1$. Функция $\varphi(\bar{X})$ называется *выпуклой вверх* (вогнутой), если $-\varphi(\bar{X})$ выпуклая вниз. При выполнении строгого неравенства функция называется *строго выпуклой* (вогнутой).

Если все $f_i(x_1, x_2, \dots, x_n)$, $i=0, 1, \dots, n$, — выпуклые вниз функции, то задача нелинейного программирования (7.143) — (7.144) является задачей выпуклого программирования.

Методы решения задач нелинейного программирования делятся на не прямые и прямые. Непрямые методы применяются при необходимости получения решения в аналитической форме, что обычно при сложных практических задачах невозможно. В этих случаях применяют прямые методы, основанные на вычислениях и сравнении функции в ряде точек. Большинство практических задач можно решать только численными методами.

Градиентные методы. Пусть необходимо определить минимум функции $f(x_1, \dots, x_n)$. Допустим выбрана произвольная точка $\bar{X}_0 = (x_1^0, \dots, x_n^0)$. В каком направлении необходимо двигаться от этой точки $\bar{X}_0$, чтобы значение функции $f(\bar{X})$, $\bar{X} = (x_1, \dots, x_n)$, уменьшилось при достаточно малом шаге, т. е. чтобы $f(\bar{X}_0) \geq f(\bar{X})$?

Предположим, что имеется такое направление движения, характеризующее n -мерным вектором $\bar{U} = (u_1, \dots, u_n)$. Согласно формуле Тейлора,

$$f(\bar{X}_0 + \rho \bar{U}) = f(\bar{X}_0) + \rho(f_x(\bar{X}_0), \bar{U}) + O(\rho), \quad (7.145)$$

где ρ — достаточно малый шаг движения от точки $\bar{X}$ в направлении $\bar{U}$; $f_x(\bar{X}_0)$ — градиент функции $f(\bar{X}_0)$, т. е. $f_x(\bar{X}) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) = \nabla f$; $O(\rho)$ — малая величина, для которой $\frac{O(\rho)}{\rho} \rightarrow 0$.

Поскольку значение функции $f(\bar{X})$ в каждой из последующих точек должно убывать и $\rho > 0$, то

$$(f_x(\bar{X}_0), \bar{U}) < 0. \quad (7.146)$$

Если направление $\bar{U}$ будет совпадать с направлением антиградиента, т. е. $\bar{U} = -f_x(\bar{X}_0)$, то движение по этому направлению приведет к точке, в которой функция $f(\bar{X})$ будет иметь меньшее значение, т. е. будет убывать.

Если точка $\bar{X} = (x_1, \dots, x_n)$ стационарна, то в ней достигается минимум (максимум) и $\frac{\partial f}{\partial \bar{X}} = 0$.

Методы последовательного улучшения значений непрерывно дифференцируемых функций при решении задач минимизации (максимизации) называются *релаксационными*.

Таким образом, решение задачи определения минимума нелинейной функции $f(\bar{X})$ сводится к отысканию точки $\bar{X}$, удовлетворяющей условию

$$\frac{\partial f}{\partial \bar{X}} = 0. \quad (7.147)$$

Если найденная точка не удовлетворяет этому условию, необходимо найти новую стационарную точку, являющуюся локальным минимумом.

Одним из релаксационных методов решения задачи минимизации более сложной функции является градиентный метод, предложенный в 1848 г. математиком Коши. Этот метод заключается в вычислении и сравнении значений функции в ряде точек, определяемых точно против направления градиента функции при задаче минимизации по правилу:

$$\bar{X}_{k+1} = \bar{X}_k - \rho_k f_x(\bar{X}_k), \quad k = 0, 1, \dots, \quad (7.148)$$

где ρ_k — шаг движения на k -й итерации; $f_x(\bar{X}_k)$ — градиент функции $f(\bar{X})$ в точке $\bar{X}_k$; $\bar{X}_0$ — произвольная точка, с которой начинается движение в окрестность точки $\bar{X}_1$ (окрестность соизмерима с шагом движения ρ_0).

Шаг движения ρ_k может быть постоянным, т. е. не изменяться от итерации к итерации, и может выбираться для каждой итерации. В первом случае метод называется *градиентным методом спуска* (движение происходит в направлении антиградиента) или *градиентным методом подъема* (движение происходит в направлении градиента), во втором случае — *методом наискорейшего спуска* (подъема). Во втором случае шаг ρ_k выбирается из условия

$$\min_{\rho \geq 0} f(\bar{X}_k - \rho f_x(\bar{X}_k)) = f(\bar{X}_k - \rho_k f_x(\bar{X}_k)). \quad (7.149)$$

По условию (7.149) шаг ρ_k определяется так, чтобы на луче, выходящем из точки $\bar{X}_k$ в направлении антиградиента ($-f_x(\bar{X}_k)$), функция $f(\bar{X})$ достигла минимального значения. Движение происходит с выбранным шагом, пока направление движения не изменится.

Для непрерывно дифференцируемой функции $f(\bar{X})$ последовательность градиентов с ростом количества итераций, т. е. при $k \rightarrow \infty$, стремится к нулю. Поэтому множество предельных точек последовательности $\{\bar{X}_k\}_k$, генерируемой в методе наискорейшего спуска, является подмножеством стационарных точек. Если количество этих точек конечно, то последовательность $\{\bar{X}_k\}$ либо сходится к одной из стационарных точек, либо имеет несколько предельных точек, причем значения функции в этих точках должны быть одинаковыми.

На каждой итерации метода наискорейшего спуска необходимо решать задачу (7.149), например методом «золотого сечения».

Существуют и другие градиентные методы отыскания экстремума функции, например методы сопряженных градиентов, метод обобщающего градиентного спуска и др.

Методы штрафных функций. Эти методы позволяют свести задачу нелинейного программирования (7.143)—(7.144) точно или приближенно к задаче на безусловный экстремум, т. е. к минимизации (максимизации) некоторой функции, представляющей сумму минимизируемой (максимизируемой) функции и некоторой функции невязок в ограничениях задачи. Метод штрафных функций впервые был предложен Курантом в 1943 г. для решения некоторых задач математической физики.

Исходя из задачи (7.143)—(7.144) строится функция

$$F(\bar{X}, \bar{s}) = f_0(\bar{X}) + S(f_1(\bar{X}), \dots, f_m(\bar{X}), \bar{s}), \quad (7.150)$$

где $\bar{s} = \{s_i\}_{i=1}^m$ — вектор управляющих параметров; S — функция штрафная (вещественная), действие которой реализуется вектором $\bar{s}$. Функция штрафа должна обладать свойством

$$\lim_{\substack{\bar{s}_i \rightarrow \infty \\ i=1, m}} S(f_1(\bar{X}), \dots, f_m(\bar{X}), \bar{s}) = \begin{cases} 0, f_i \leq 0, i = \overline{1, m}; \\ +\infty — \text{в противном случае.} \end{cases} \quad (7.151)$$

Идея метода штрафных функций сводится к минимизации функции $F(\bar{X}, \bar{s})$ вместо решения задачи (7.143)—(7.144). Штрафные функции могут быть заданы в различном виде, определены на всем пространстве, но в области допустимых решений они равны нулю, вне этой области положительны и возрастают при $f_i(\bar{X}) \rightarrow \infty, i = \overline{1, m}$.

Примером функции штрафа может быть функция

$$S(f, s) = \frac{1}{s} S(\bar{X}) = \frac{1}{s} \sum_{i=1}^m S^i(f_i(\bar{X})),$$

где s — положительное число; $S^i(f_i(\bar{X}))$ — определенная при любом $f_i(\bar{X})$ непрерывная функция; $S^i(f_i(\bar{X})) \rightarrow +\infty$ при $f_i(\bar{X}) \rightarrow +\infty$ и $S^i(f_i(\bar{X})) = 0$ при $f_i(\bar{X}) \leq 0$ для ограничений типа неравенств или $S^i(f_i(\bar{X})) \rightarrow +\infty$ при $f_i(\bar{X}) \rightarrow +\infty$ для ограничений типа равенств.

В качестве функций штрафа могут быть также следующие функции:

Для ограничения исходной задачи типа неравенств

$$S^i(f_i(\bar{X})) = [\max \{0, f_i(\bar{X})\}]^2;$$

Для ограничений исходной задачи типа равенств

$$S^i(f_i(\bar{X})) = (f_i(\bar{X}))^2.$$

Обычно вид функции штрафа выбирается в зависимости от требования методов, применяемых для минимизации функции $F(\bar{X}, \bar{s})$.

Рассмотрим решение задач нелинейного программирования в случае ограничений-равенств. Необходимо решить задачу: минимизировать

$$\min f_0(x_1, x_2, \dots, x_n) \quad (7.152)$$

при ограничениях

$$f_i(x_1, x_2, \dots, x_n) = 0, \quad i = \overline{1, m}. \quad (7.153)$$

Пусть целевая функция и функции-ограничения $f_i(x_1, \dots, x_n)$, $i = \overline{0, 1, \dots, m}$, дифференцируемы.

Вводятся множители Лагранжа λ_i , $i = \overline{1, m}$, и строится функция Лагранжа

$$L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f_0(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i f_i(x_1, \dots, x_n). \quad (7.154)$$

Вектор $\bar{X}_0 = (x_1^0, \dots, x_n^0)$ является оптимальным решением задачи (7.152) — (7.153), если существует вектор $\bar{\lambda}_0 = (\lambda_1^0, \dots, \lambda_m^0)$, при котором выполняются соотношения:

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} = \frac{\partial f_0(\bar{X}_0)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial f_i(\bar{X}_0)}{\partial x_j} = 0, \quad j = \overline{1, n}; \quad (7.155)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} = f_i(\bar{X}_0) = 0, \quad i = \overline{1, m}. \quad (7.156)$$

Система (7.155) — (7.156) содержит $n + m$ уравнений и столько же неизвестных x_j , $j = \overline{1, n}$, и λ_i , $i = \overline{1, m}$.

Решая систему (7.155) — (7.156), находят решения задачи (7.152) — (7.153), т. е. $\bar{X}_0 = (x_1^0, x_2^0, \dots, x_n^0)$, и исследуют эти точки на экстремум.

Рассмотрим решение задачи нелинейного программирования в случае ограничений-неравенств. Пусть необходимо решить задачу:

минимизировать

$$f_0(x_1, \dots, x_n) \quad (7.157)$$

при ограничениях

$$f_i(x_1, \dots, x_n) \leq 0, \quad i = \overline{1, m}. \quad (7.158)$$

Предполагается, что целевая функция $f_0(x_1, \dots, x_n)$ нелинейна, а ограничения $f_i(x_1, \dots, x_n)$, $i = \overline{1, m}$, могут быть как линейными, так и нелинейными функциями. Если точка $\bar{X}_0$ является решением задачи (7.157) — (7.158), среди ограничений (7.158) найдутся такие, что $f_r(\bar{X}_0) = 0$ и $f_l(\bar{X}_0) < 0$, $r \neq l$.

Допустим, что ограничения (7.158) являются линейными функциями, т. е. имеют вид

$$f_i(x_1, \dots, x_n) = -\bar{B}_i \bar{X} + s_i, \quad i = \overline{1, m}. \quad (7.159)$$

В этом случае область допустимых решений является выпуклым множеством (многогранником) D и вместо задачи на условный

экстремум (7.157) — (7.159) можно решить задачу на безусловный экстремум функции

$$F(x_1, \dots, x_n) = f_0(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i f_i(x_1, \dots, x_n), \lambda_i \geq 0, \quad (7.160)$$

причем оптимальное решение $\bar{X}_0 = (x_1^0, \dots, x_n^0)$ удовлетворяет следующим условиям:

$$\overline{\nabla F}(x_1^0, \dots, x_n^0) = \overline{\nabla f_0}(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i \overline{\nabla f_i}(x_1, \dots, x_n) = 0; \quad (7.161)$$

$$\sum_{i=1}^m \lambda_i f_i(x_1^0, \dots, x_n^0) = 0, \lambda_i \geq 0, i = \overline{1, m}. \quad (7.162)$$

В случае ограничений вида неравенств ($f_i(x_1, \dots, x_n) \leq 0$) удовлетворение соотношений (7.161) — (7.162) требует условия регулярности ограничений Слейтера, т. е. существует по крайней мере одна точка $\bar{X}$, для которой $f_i(\bar{X}) \leq 0, i = \overline{1, m}$.

Условия (7.161), (7.162) являются условиями оптимальности задачи нелинейного программирования в случае линейных ограничений.

Условия оптимальности задачи (7.157) — (7.158) определяются основной теоремой нелинейного программирования — теоремой Куна — Таккера.

Пусть целевая функция и ограничения $f_i(\bar{X}), i = 0, 1, \dots, m$, имеют непрерывные частные производные на некотором множестве, содержащем точку $\bar{X}_0$. Если эта точка является решением задачи (7.157) — (7.158) и ограничения удовлетворяют условиям регулярности в виде линейной независимости, то существуют такие множители Лагранжа $\lambda_i, i = \overline{1, m}$, при которых выполняются условия:

$$\overline{\nabla f_0}(\bar{X}_0) + \sum_{i=1}^m \lambda_i \overline{\nabla f_i}(\bar{X}_0) = 0; \quad (7.163)$$

$$\sum_{i=1}^m \lambda_i f_i(\bar{X}_0) = 0, \lambda_i \geq 0, i = \overline{1, m}. \quad (7.164)$$

Введя функцию Лагранжа

$$L(\bar{X}, \bar{\lambda}) = f_0(\bar{X}) + \sum_{i=1}^m \lambda_i f_i(\bar{X}), \quad (7.165)$$

условия (7.163) — (7.164) можно переписать в виде

$$\overline{\nabla_x L}(\bar{X}, \bar{\lambda}) = 0; \quad (7.166)$$

$$\overline{\nabla_\lambda L}(\bar{X}, \bar{\lambda}) \leq 0; \quad (7.167)$$

$$\bar{\lambda}^T \overline{\nabla_\lambda L}(\bar{X}, \bar{\lambda}) = \sum_{i=1}^m \lambda_i f_i(\bar{X}) = 0. \quad (7.168)$$

Если целевая функция и все ограничения $f_i(\bar{X}), i = 0, 1, \dots, m$, выпуклы и ограничения $f_i(\bar{X}), i = \overline{1, m}$, удовлетворяют условию регулярности Слейтера, то вектор $\bar{X}_0$ будет решением задачи (7.157) — (7.158) тогда и только тогда, когда существует вектор

$\bar{\lambda}_0 \geq 0$, при котором пара векторов $\bar{X}_0, \bar{\lambda}_0$ является седловой точкой, т. е. выполняются условия:

$$L(\bar{X}_0, \bar{\lambda}) \leq L(\bar{X}_0, \bar{\lambda}_0) \leq L(\bar{X}, \bar{\lambda}_0); \quad (7.169)$$

$$\lambda_i^0 f_i(\bar{X}_0) = 0, \quad i = \overline{1, m}. \quad (7.170)$$

В седловой точке $(\bar{X}_0, \bar{\lambda}_0)$

$$L(\bar{X}_0, \bar{\lambda}_0) = \max_{\bar{\lambda} \geq 0} \min_{\bar{X} \in R^n} L(\bar{X}, \bar{\lambda}) = \min_{\bar{X} \in R^n} \max_{\bar{\lambda} \geq 0} L(\bar{X}, \bar{\lambda}). \quad (7.171)$$

Следовательно, решение задачи нелинейного программирования сводится к отысканию седловой точки функции Лагранжа.

Если в задаче нелинейного программирования на вектор управляющих воздействий накладывается условие неотрицательности, т. е. $\bar{X}_0 \geq 0$, а функция цели и ограничения $f_i(\bar{X})$, $i = \overline{0, 1, \dots, m}$, дифференцируемы и выпуклы по $\bar{X}$, то вектор $\bar{X}_0 \geq 0$ будет оптимальным решением задачи тогда и только тогда, когда существует такой вектор $\bar{\lambda}_0 \geq 0$, что пара векторов $\bar{X}_0, \bar{\lambda}_0$ является седловой точкой функции Лагранжа $L(\bar{X}, \bar{\lambda})$, т. е. выполняются следующие условия:

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} \geq 0, \quad j = \overline{1, n}; \quad (7.172)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} x_j^0 = 0, \quad j = \overline{1, n}; \quad (7.173)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} = f_i(\bar{X}_0) \leq 0, \quad i = \overline{1, m}; \quad (7.174)$$

$$\lambda_i^0 f_i(\bar{X}_0) = 0, \quad i = \overline{1, m}. \quad (7.175)$$

Для задачи нелинейного программирования максимизировать

$$f_0(x_1, \dots, x_n) \quad (7.176)$$

при ограничениях

$$f_i(x_1, \dots, x_n) \geq 0, \quad i = \overline{1, m}; \quad (7.177)$$

$$\bar{X} = (x_1, \dots, x_n) \geq 0 \quad (7.178)$$

условия оптимальности следующие:

если $f_i(\bar{X})$, $i = \overline{0, 1, \dots, m}$, — дифференцируемые и вогнутые функции, то для того, чтобы вектор $\bar{X}_0$ являлся оптимальным решением задачи (7.176) — (7.178), необходимо и достаточно, чтобы существовал такой вектор $\bar{\lambda}_0 \geq 0$, для которого выполняются условия:

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} \leq 0, \quad j = \overline{1, n}; \quad (7.179)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} x_j^0 = 0, \quad j = \overline{1, n}; \quad (7.180)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} = f_i(\bar{X}_0) \geq 0, \quad i = \overline{1, m}; \quad (7.181)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} \lambda_i^0 = 0, \quad i = \overline{1, m}. \quad (7.182)$$

Метод Эрроу — Гурвица. Пусть задача нелинейного программирования имеет вид:
максимизировать

$$f_0(x_1, \dots, x_n) \quad (7.183)$$

при ограничениях

$$f_i(x_1, \dots, x_n) \geq 0, \quad i = \overline{1, m}; \quad (7.184)$$

$$\bar{X} = (x_1, \dots, x_n) \geq 0. \quad (7.185)$$

Если все функции $f_i(\bar{X})$, $i = 0, 1, \dots, m$, — дифференцируемы и выпуклы вверх, то решение задачи (7.183) — (7.185) сводится к отысканию седловой точки функции Лагранжа

$$L(\bar{X}, \bar{\lambda}) = f_0(\bar{X}) + \sum_{i=1}^m \lambda_i f_i(\bar{X}),$$

т. е. к решению задачи $\max \min L(\bar{X}, \bar{\lambda})$.
 $\bar{X} \geq 0 \quad \bar{\lambda} \geq 0$

Седловую точку $(\bar{X}_0, \bar{\lambda}_0)$ можно найти по правилу

$$\bar{X}_{k+1} = \bar{X}_k + \rho \overline{\nabla_x L}(\bar{X}_k, \bar{\lambda}_k); \quad (7.186)$$

$$\bar{\lambda}_{k+1} = \max \{0, \bar{\lambda}_k - \rho \overline{\nabla_\lambda L}(\bar{X}_k, \bar{\lambda}_k)\}, \quad (7.187)$$

где

$$\overline{\nabla_x L}(\bar{X}_k, \bar{\lambda}_k) = \left\{ \frac{\partial L(\bar{X}_k, \bar{\lambda}_k)}{\partial x_1}, \dots, \frac{\partial L(\bar{X}_k, \bar{\lambda}_k)}{\partial x_n} \right\}; \quad (7.188)$$

$$\overline{\nabla_\lambda L}(\bar{X}_k, \bar{\lambda}_k) = \left\{ \frac{\partial L(\bar{X}_k, \bar{\lambda}_k)}{\partial \lambda_1}, \dots, \frac{\partial L(\bar{X}_k, \bar{\lambda}_k)}{\partial \lambda_m} \right\}. \quad (7.189)$$

Таким образом, для отыскания седловой точки $(\bar{X}_0, \bar{\lambda}_0)$ необходимо осуществить движение по переменным x_j , $j = \overline{1, n}$, в направлении градиента функции $L(\bar{X}, \bar{\lambda})$, а по переменным λ_i , $i = \overline{1, m}$, — по антиградиенту этой функции. Последовательность $\bar{X}_k$, $\bar{\lambda}_k$ сходится к седловой точке, если выполняются условия Слейтера

$$\sum_{i=1}^m \lambda_i f_i(x_1, \dots, x_n) = 0.$$

Квадратичное программирование. Пусть необходимо решить задачу:

максимизировать

$$\sum_{j=1}^n c_j x_j + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n c_{jk} x_j x_k \quad (7.190)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j + r_i = a_{i0}, \quad i = \overline{1, m}; \quad (7.191)$$

$$x_j \geq 0, \quad j = \overline{1, n}; \quad (7.192)$$

$$r_i \geq 0, \quad i = \overline{1, m}. \quad (7.193)$$

Таким образом, функция цели $f_0(\bar{X})$, имеющая вид (7.190), является квадратичной, вогнутой, а все ограничения — линейные.

Раздел математического программирования, рассматривающий специальный класс задач, в которых максимизируемая (минимизируемая) функция квадратична, а ограничения линейны, получил название «квадратичного программирования».

Условия теоремы Куна — Таккера позволяют перейти от нелинейной задачи (7.190) — (7.193) к линейной.

В векторном виде задача (7.190) — (7.193) записывается так: максимизировать

$$\left(\bar{B}^T \bar{X} + \frac{1}{2} \bar{X}^T \bar{C} \bar{X} \right) \quad (7.194)$$

при ограничениях

$$\bar{A}_0 - \bar{A} \bar{X} \geq 0; \quad (7.195)$$

$$\bar{X} \geq 0. \quad (7.196)$$

Условия теоремы Кунна — Таккера для задачи (7.194) — (7.196) следующие:

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \bar{X}} = \bar{B} + \bar{C} \bar{X} - \bar{A}^T \bar{\lambda} + \bar{V} = 0; \quad (7.197)$$

$$\frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \bar{\lambda}} = \bar{A}_0 - \bar{A} \bar{X} - \bar{W} = 0; \quad (7.198)$$

$$\bar{V}^T \bar{X} = 0; \quad (7.199)$$

$$\bar{W}^T \bar{\lambda} = 0. \quad (7.200)$$

Здесь функция Лагранжа имеет вид

$$L(\bar{X}, \bar{\lambda}) = \bar{B}^T \bar{X} + \frac{1}{2} \bar{X}^T \bar{C} \bar{X} + \bar{\lambda}^T (\bar{A}_0 - \bar{A} \bar{X});$$

$\bar{V}^T = (v_1, \dots, v_n)$ — вектор остаточных переменных; $\bar{W}^T = (w_1, \dots, w_m)$ — вектор избыточных переменных.

$$v_j > 0, \text{ если } \frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} < 0 \text{ и } v_j = 0, \text{ если } \frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial x_j} = 0;$$

$$w_i > 0, \text{ если } \frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} > 0 \text{ и } w_i = 0, \text{ если } \frac{\partial L(\bar{X}_0, \bar{\lambda}_0)}{\partial \lambda_i} = 0.$$

Полученные первые два условия образуют систему из $n+m$ уравнений с $2(n+m)$ неизвестными $\bar{X}, \bar{\lambda}, \bar{V}, \bar{W}$. Последние два условия (условия Слейтера) свидетельствуют о том, что n переменных из $\bar{X}$ и $\bar{V}$ и m переменных из $\bar{\lambda}$ и $\bar{W}$ должны быть равны нулю. Поэтому, если существует решение задачи (7.194) — (7.196), то оно должно быть одним из базисных решений задачи:

$$\bar{C}\bar{X} - \bar{A}^T\bar{\lambda} + \bar{V} = -\bar{B}; \quad (7.201)$$

$$\bar{A}\bar{X} + \bar{W} = \bar{A}_0. \quad (7.202)$$

Применяя метод искусственных переменных, получим задачу: минимизировать

$$\left(\sum_{i=1}^m My_i + \sum_{j=1}^n Mz_j \right) \quad (7.203)$$

при ограничениях

$$\bar{C}\bar{X} - \bar{A}^T\bar{\lambda} + \bar{V} + \bar{Z} = -\bar{B}; \quad (7.204)$$

$$\bar{A}\bar{X} + \bar{W} + \bar{Y} = \bar{A}_0, \quad (7.205)$$

где

$$\bar{Z} = (z_1, z_2, \dots, z_n); \quad \bar{Y} = (y_1, y_2, \dots, y_m);$$

$$\bar{Z} \geq 0; \quad \bar{Y} \geq 0.$$

При решении задачи (7.203) — (7.205) необходимо учитывать условия Слейтера: $\bar{V}^T\bar{X} = 0$ и $\bar{W}^T\bar{\lambda} = 0$.

Если удастся вывести все искусственные переменные, то найденное базисное решение будет оптимальным.

Пример 1. Решить задачу:

минимизировать

$$f(x_1, x_2) = (x_1 - 4)^2 + (x_2 - 3)^2$$

при ограничении

$$x_1 + x_2 = 9.$$

Решение. Составляем функцию Лагранжа

$$L(x_1, x_2, \lambda) = (x_1 - 4)^2 + (x_2 - 3)^2 + \lambda(9 - x_1 - x_2).$$

Определяем частные производные:

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_1} = 2(x_1 - 4) - \lambda = 0;$$

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_2} = 2(x_2 - 3) - \lambda = 0;$$

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial \lambda} = 9 - x_1 - x_2 = 0.$$

Решаем систему и находим: $x_1^0 = 5$; $x_2^0 = 4$; $\lambda = 2$.

Исследуем функцию $f(x_1, x_2) = (x_1 - 4)^2 + (x_2 - 3)^2$ в окрестности точки

$$\bar{X}_0 = (x_1^0 = 5, x_2^0 = 4):$$

$$f_{11} = \frac{\partial^2 f(\bar{X}_0)}{\partial x_1^2} = 2 > 0; \quad f_{22} = \frac{\partial^2 f(\bar{X}_0)}{\partial x_2^2} = 2 > 0;$$

$$\frac{\partial^2 f(\bar{X}_0)}{\partial x_1 \partial x_2} = f_{12} = f_{21} = 0.$$

Поскольку $f_{11}(\bar{X}_0) > 0$ и $\begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} > 0$, то функция $f(x_1, x_2)$ выпукла и в точ-

ке $\bar{X}_0 = (5, 4)$ имеет абсолютный минимум.

Пример 2. Решить задачу:
максимизировать

$$f(x_1, x_2) = (44x_1 + 130x_2 - 4x_1^2 - 16x_2^2)$$

при ограничениях

$$4x_1 + 7x_2 \leq 30; \quad 4x_1 - x_2 \leq 10; \quad x_1 \geq 0; \quad x_2 \geq 0.$$

Решение. Исследуем целевую функцию:

$$f_1 = \frac{\partial f}{\partial x_1} = 44 - 8x_1; \quad f_{11} = \frac{\partial^2 f}{\partial x_1^2} = -8 < 0;$$

$$f_2 = \frac{\partial f}{\partial x_2} = 130 - 32x_2; \quad f_{22} = \frac{\partial^2 f}{\partial x_2^2} = -32 < 0;$$

$$f_{12} = f_{21} = \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0; \quad \Delta = \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = \begin{vmatrix} -8 & 0 \\ 0 & -32 \end{vmatrix} > 0.$$

Поскольку $f_{11} < 0$, а $\Delta > 0$, то функция $f(x_1, x_2)$ вогнута и задача является задачей квадратичного программирования.

Составляем функцию Лагранжа

$$L(x_1, x_2, \lambda_1, \lambda_2) = (44x_1 + 130x_2 - 4x_1^2 - 16x_2^2) + \\ + \lambda_1(30 - (4x_1 + 7x_2)) + \lambda_2(10 - (4x_1 - x_2)).$$

Применим теорему Куна — Таккера и получим условия для определения оптимального решения:

$$\left. \begin{aligned} \frac{\partial L}{\partial x_1} &= 44 - 8x_1 - 4\lambda_1 - 4\lambda_2 \leq 0; \\ \frac{\partial L}{\partial x_2} &= 130 - 32x_2 - 7\lambda_1 + \lambda_2 \leq 0; \\ \frac{\partial L}{\partial \lambda_1} &= 30 - 4x_1 - 7x_2 \geq 0; \\ \frac{\partial L}{\partial \lambda_2} &= 10 - 4x_1 + x_2 \geq 0; \end{aligned} \right\} \text{I} \quad \left. \begin{aligned} \frac{\partial L}{\partial x_1} x_1 &= 0; \\ \frac{\partial L}{\partial x_2} x_2 &= 0; \\ \frac{\partial L}{\partial \lambda_1} \lambda_1 &= 0; \\ \frac{\partial L}{\partial \lambda_2} \lambda_2 &= 0. \end{aligned} \right\} \text{II}$$

Вводя в эти системы свободные переменные (остаточные и избыточные) v_1, v_2 и w_1, w_2 , получим систему и условия дополняющей нежесткости:

$$\left. \begin{aligned} 44 - 8x_1 - 4\lambda_1 - 4\lambda_2 + v_1 &= 0; \\ 130 - 32x_2 - 7\lambda_1 + \lambda_2 + v_2 &= 0; \\ 30 - 4x_1 - 7x_2 - w_1 &= 0; \\ 10 - 4x_1 + x_2 - w_2 &= 0; \end{aligned} \right\} \text{III} \quad \left. \begin{aligned} x_1 v_1 &= 0; \\ x_2 v_2 &= 0; \\ \lambda_1 w_1 &= 0; \\ \lambda_2 w_2 &= 0. \end{aligned} \right\} \text{IV}$$

Перепишем систему в виде:

$$\left. \begin{aligned} 8x_1 + 4\lambda_1 + 4\lambda_2 - v_1 &= 44; \\ 32x_2 + 7\lambda_1 - \lambda_2 - v_2 &= 130; \\ 4x_1 + 7x_2 + w_1 &= 30; \\ 4x_1 - x_2 + w_2 &= 10. \end{aligned} \right\} V$$

Для решения задачи необходимо найти допустимое базисное решение системы V при выполнении условий IV.

Решение можно получить, применив симплекс-метод, для чего необходимо ввести искусственные переменные, так как в системе V нет единичного базиса. Введя искусственные переменные в первое и второе уравнения системы V, получим следующую задачу линейного программирования:

$$\text{минимизировать} \quad M(y_1 + y_2)$$

при ограничениях

$$\left. \begin{aligned} 8x_1 + 4\lambda_1 + 4\lambda_2 - v_1 + y_1 &= 44; \\ 32x_2 + 7\lambda_1 - \lambda_2 - v_2 + y_2 &= 130; \\ 4x_1 + 7x_2 + w_1 &= 30; \\ 4x_1 - x_2 + w_2 &= 10. \end{aligned} \right\} VI$$

Таким образом, задача квадратичного программирования свелась к задаче линейного программирования.

7.5. СТОХАСТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Стохастическое программирование предназначено для решения задач стохастической природы, т. е. задач, характеризуемых случайными величинами. Особенности таких задач связаны с отсутствием полной информации о функции цели и функциях ограничений, об их производных, а также с негладким характером этих функций. Часто в таких задачах нельзя для заданного решения определить точное значение функции цели и проверить выполнение заданных ограничений.

Термин «стохастическое программирование» введен в начале 50-х годов математиками Данцигом, Чернсом, Купером при изучении задач линейного программирования со случайными коэффициентами. Обычно эти задачи гораздо сложнее задач нелинейного программирования и для их решения разработаны специальные численные методы, обобщающие детерминированные методы нелинейного программирования.

Стохастическое программирование включает методы решения задач в условиях риска и неопределенности. Задачи в условиях риска характеризуются известными вероятностями возможных частных исходов принимаемых решений.

Для задач в условиях неопределенности вероятности возможных исходов принимаемых решений неизвестны.

Пусть имеются m допустимых решений, n возможных исходов этих решений и c_{ij} — затраты, связанные с решением i , $i = \overline{1, m}$, при исходе j , $j = \overline{1, n}$. Истинный исход при принятии решения i неизвестен. Задача заключается в выборе такого допустимого решения i , которое в смысле некоторого критерия является наилучшим.

Если бы любое решение однозначно приводило к определенному исходу, то задача была бы эквивалентна задачам линейного или нелинейного программирования, так как необходимо выбрать такое решение i при заданном исходе j , для которого величина c_{ij} была бы минимальной. Но исход j не задан, потому задача выбора решения усложняется.

При решении задачи в условиях риска заданы вероятности исходов $p_1, \dots, p_n$, причем $\sum_{j=1}^n p_j = 1$. Часто критерием выбора решения i является минимизация средних затрат

$$f_i = \sum_{j=1}^n c_{ij} p_j, \quad i = \overline{1, m}, \quad (7.206)$$

или максимизация вероятности

$$f_i = P\{c_{ij} \leq c, j = \overline{1, n}\}, \quad i = \overline{1, m}, \quad (7.207)$$

где c — некоторые заданные затраты.

Критерии оптимальности необходимо выбирать в зависимости от конкретной ситуации, в которой принимается решение.

При решении задач в условиях неопределенности, т. е. в условиях, когда вероятности исходов $p_j, j = \overline{1, n}$, неизвестны, может быть принят один из следующих критериев оптимальности решения.

1. Минимизация максимально возможного убытка (гарантированный уровень)

$$f = \min_{1 \leq i \leq m} \max_{1 \leq j \leq n} c_{ij}. \quad (7.208)$$

Оптимальным является то решение $i_0 = i$, при котором максимальный убыток достигает минимума.

2. Критерий минимаксного риска (критерий Сэвиджа)

$$f = \min_{1 \leq i \leq m} \max_{1 \leq j \leq n} \bar{c}_{ij} = \min_{1 \leq i \leq m} \max_{1 \leq j \leq n} (c_{ij} - \min_{1 \leq i \leq m} c_{ij}), \quad (7.209)$$

где $\bar{c}_{ij}$ — элементы матрицы риска

$$\bar{C} = \|\bar{c}_{ij}\|_{m \times n} = \|c_{ij} - \min_{1 \leq i \leq m} c_{ij}\|_{m \times n}.$$

3. Критерий оптимальности по Парето. Решение i называется оптимальным, если не существует $s \neq i$, при котором

$$c_{sj} \leq c_{ij}, \quad j = \overline{1, n}. \quad (7.210)$$

4. Принцип недостаточного основания Бернулли (принцип равновероятности исходов). Если нет оснований считать, что какой-то исход вероятней другого исхода, то все исходы принимаются равновероятными. Этим задачу в условиях неопределенности сводят к задаче в условиях риска.

На практике часто вероятности исходов определяются на основе экспертных оценок.

Модель задачи стохастического программирования имеет вид:
минимизировать

$$f_0(\bar{X}, \omega) \quad (7.211)$$

при ограничениях

$$f_i(\bar{X}, \omega) \leq 0, \quad i = \overline{1, m}; \quad (7.212)$$

$$\bar{X} \in X, \quad (7.213)$$

где ω — элемент некоторого вероятностного пространства; X — множество решений.

С каждым решением $\bar{X}$ можно связать числовые параметры $f_t(\bar{X}, \omega)$, $t = 0, 1, \dots, m$, зависящие от него и от случайных параметров ω . Решения $\bar{X}$ в одних задачах могут быть детерминированными, а в других — зависящими от случайного параметра ω .

Различают задачи оперативного и задачи перспективного стохастического программирования.

Для задач оперативного стохастического программирования характерна возможность уточнения случайного параметра ω до выбора решения $\bar{X}$; это решение зависит от ω и является случайным вектором. Так как эксперимент для уточнения параметра ω полностью его не определяет, то наряду с пространством состояний ω имеется вероятностное пространство наблюдений. Поэтому решение $\bar{X}$ должно быть измеримой функцией пространства наблюдений. Для упрощения принимают, что пространство наблюдений совпадает с пространством состояний и не зависит от решений $\bar{X}$. Такая постановка задачи характерна при оперативном текущем технико-экономическом планировании, диагностике и т. п.

Задачи перспективного стохастического программирования характерны тем, что решение $\bar{X}$ принимается до экспериментов для уточнения случайного параметра ω и детерминированным. Примером таких задач могут быть задачи перспективного технико-экономического планирования, задачи проектирования и др. В таких задачах параметры изучаемой системы должны быть детерминированными величинами в определенном диапазоне возмущений. В этих задачах нет очень жестких требований к правилам выбора решений; существенным является фактор неопределенности, что часто приводит к недопустимости детерминированных решений.

В задачах оперативного стохастического программирования выбор решения обычно происходит в реальном масштабе времени, поэтому рассматриваются только простые правила выбора решений $\bar{X}$.

В задачах перспективного стохастического программирования при детерминированном решении $\bar{X}$ модель (7.211) — (7.213) часто заменяется следующей моделью:

минимизировать

$$F_0(\bar{X}) = Ef_0(\bar{X}, \omega) = \int f_0(\bar{X}, \omega) P(d\omega) \quad (7.214)$$

при ограничениях

$$F_i(\bar{X}) = Ef_i(\bar{X}, \omega) = \int f_i(\bar{X}, \omega) P(d\omega); \quad i = \overline{1, m}; \quad (7.215)$$

$$\bar{X} \in X \quad (7.216)$$

или минимизировать

$$\Psi_0(\bar{X}) = P\{f_0(\bar{X}, \omega) \geq a\} \quad (7.217)$$

при ограничениях

$$\Psi_i(\bar{X}) = P\{f_i(\bar{X}, \omega) \leq 0\} - p_i \geq 0, \quad i = \overline{1, m}; \quad (7.218)$$

$$\bar{X} \in X, \quad (7.219)$$

где $F_\xi(\bar{X})$ — математическое ожидание функции $f_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$; $\Psi_\xi(\bar{X})$ — вероятность отклонения функции $f_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$; a, p_ξ — некоторые константы.

Могут рассматриваться модели (7.214), (7.218), (7.219) или (7.217), (7.215), (7.216).

Рассмотренные модели могут быть дополнены дисперсиями функций $f_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$, или другими моментами этих функций.

В задачах оперативного стохастического программирования решение $\bar{X}$ зависит от случайного параметра ω , т. е. является случайным. В результате эксперимента из всей совокупности событий, образующих σ -алгебру F основного вероятностного пространства, доступна наблюдению некоторая часть событий некоторой σ -подалгебры G , $G \subseteq F$. Поэтому функция $\bar{X}(\omega)$ должна быть измеримой относительно G (G -измеримой).

В этих условиях модель задачи стохастического программирования имеет вид:

минимизировать

$$F_0(\bar{X}(\omega)) = Ef_0(\bar{X}(\omega), \omega) \quad (7.220)$$

при ограничениях

$$F_i(\bar{X}(\omega)) = Ef_i(\bar{X}(\omega), \omega) \leq 0, \quad i = \overline{1, m}; \quad (7.221)$$

$$\bar{X} \in X. \quad (7.222)$$

Таким образом, задача стохастического программирования заключается в том, что необходимо найти G -измеримую вектор-функцию $\bar{X}(\omega) = (x_1(\omega), x_2(\omega), \dots, x_n(\omega))$, которая удовлетворяет ограничениям (7.221) — (7.222) и минимизирует целевую функцию (7.220).

Задачи оперативного стохастического программирования могут быть сведены к задачам перспективного стохастического программирования введением некоторых решающих правил.

Рассмотрим методы решения задач стохастического программирования.

В задачах, описываемых моделью (7.214) — (7.216), обычно имеется возможность определить в результате эксперимента значения функций $f_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$, для фиксированных ω . Если математические ожидания $F_\xi(\bar{X})$, $\xi = 0, 1, \dots, m$, можно определить, задача (7.214) — (7.216) сводится к задаче нелинейного программирования, так как стохастическая природа такой задачи проявляется только при определении математических ожиданий функций $f_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$, т. е. при определении $F_\xi(\bar{X}, \omega)$, $\xi = 0, 1, \dots, m$.

Часто стремятся определить $F_i(\bar{X}, \omega)$, $i=0, 1, \dots, m$, а затем, применяя методы нелинейного программирования, решить задачу стохастического программирования (7.214) — (7.216).

В случае невозможности определения $F_i(\bar{X}, \omega)$, $i=0, 1, \dots, m$, задачу (7.214) — (7.216) заменяют приближенным детерминированным эквивалентом, полученным при замене случайной величины ω его средним значением $\bar{\omega}$.

Методы решений, основанные на информации о значениях $f_i(\bar{X}, \omega)$, $i=0, 1, \dots, m$, называют прямыми методами стохастического программирования, все остальные методы — непрямыми.

Методы, основанные на сведении задачи к некоторой другой, эквивалентной задаче, также называют непрямыми.

Основой не прямых методов стохастического программирования является применение необходимых признаков экстремума, сведение или подмена стохастической экстремальной задачи детерминированным аналогом — задачей нелинейного программирования. Разработано много специальных приемов не прямых методов, хотя ими решается узкий класс задач.

Если в задаче (7.214) — (7.216) вектор $\bar{X}$ — детерминированный (задача перспективного стохастического программирования), то эта задача формально является задачей нелинейного программирования. Если вектор $\bar{X}$ — функция G -измеримая относительно исходного вероятностного пространства (Ω, F, P) (задача оперативного стохастического программирования), то задача (7.214) — (7.216) является задачей программирования в абстрактных пространствах. Полученные условия оптимальности в этих случаях нельзя проверить на оптимальность, так как производные функций $F_i(\bar{X})$ неизвестны.

В практических задачах решение $\bar{X}(\omega)$ кроме измеримости относительно некоторой σ -подалгебры решений должно удовлетворять определенным требованиям реализуемости в рассматриваемой системе принятия решений. Например, если $\bar{X}(\omega)$ — управляющее воздействие в системе управления движением объекта и ω — отклонение этого объекта от расчетной траектории, то $\bar{X}(\omega)$ можно отыскивать не среди всевозможных функций от ω , а только среди линейных функций $\varphi(\bar{Y}, \omega)$ (где $\bar{Y}$ — неизвестные параметры), так как принцип управления пропорционально отклонению легко реализуется.

Отыскание вектора $\bar{X}(\omega)$ в классе измеримых функций связано с большими временными затратами, а решение задачи часто необходимо внедрять быстро, непосредственно за наблюдением. Поэтому прибегают к параметризации решения $\bar{X}(\omega)$, т. е. решение $\bar{X}(\omega)$ отыскивается в виде наперед заданной функции $\bar{X}(\omega) = \bar{X}(\bar{Y}, \omega) = [x_1(\bar{Y}, \omega), \dots, x_n(\bar{Y}, \omega)]$, определенной с точностью до вектора $\bar{Y} = (y_1, \dots, y_l)$, значение которого вычисляется в зависимости от распределения $P(d\omega)$ до наблюдения ω . В результате такой параметризации задача оперативного стохастического программирования сводится к задаче перспективного стохастического программирования.

Рассмотрим частные случаи.

Допустим, что пространство наблюдений совпадает с пространством состояний, т. е. в результате наблюдений ω становится известным. В этом случае $\bar{X}(\omega)$ является решением задачи: минимизировать

$$f_0(\bar{X}, \omega) \quad (7.223)$$

при ограничениях

$$f_i(\bar{X}, \omega) \leq 0, \quad i = \overline{1, m}; \quad (7.224)$$

$$\bar{X} \in X. \quad (7.225)$$

Часто возникает задача, как определить решение $\bar{X}(\omega)$ при новом значении ω , не прибегая к решению задачи (7.223) — (7.225), а используя решение при старом ω .

1. Пусть задача (7.223) — (7.225) является задачей линейного программирования при заданном ω : минимизировать

$$\bar{C}^T \bar{X} \quad (7.226)$$

при ограничениях

$$\bar{A} \bar{X} \leq \bar{A}_0(\omega); \quad (7.227)$$

$$\bar{X} \geq 0. \quad (7.228)$$

В этом случае $\bar{X}(\omega) = \bar{Z} \bar{A}_0(\omega)$, где $\bar{Z}$ — детерминированная матрица с неизвестными элементами, значения которых определяются до наблюдения $\bar{A}_0(\omega)$. Матрицу $\bar{Z}$ можно выбрать из условия минимума

$$F_0(\bar{Z}) = P\{\bar{C}, \bar{Z} \bar{A}_0(\omega)\} \geq a \quad (7.229)$$

при ограничениях

$$F_i(\bar{Z}) = P\{(\bar{A} \bar{Z} \bar{A}_0(\omega))_i \leq a_{i0}(\omega)\} \geq p_i, \quad i = \overline{1, m}, \quad (7.230)$$

где $a, p_i, i = \overline{1, m}$ — некоторые константы.

2. Пусть имеется задача стохастического программирования с линейными ограничениями: минимизировать

$$\sum_{j=1}^n c_j(\omega) x_j \quad (7.231)$$

при ограничениях

$$\sum_{j=1}^n a_{ij}(\omega) x_j \leq a_{i0}(\omega), \quad i = \overline{1, m}; \quad (7.232)$$

$$x_j \geq 0, \quad j = \overline{1, n}. \quad (7.233)$$

Решение $\bar{X} = (x_1, \dots, x_n)$ выбирается после наблюдения ω . Пусть z_{ij} удовлетворяют условиям:

$$\sum_{j=1}^n z_{ij} = 1, \quad 0 \leq z_{ij} \leq 1, \quad i = \overline{1, m}; \quad j = \overline{1, n}. \quad (7.234)$$

Задачу (7.231) — (7.233) можно свести к n задачам минимизации:

минимизировать

$$c_j(\omega) x_j \quad (7.235)$$

при ограничениях

$$a_{ij}(\omega) x_j \leq a_{i0}(\omega) z_{ij}, \quad i = \overline{1, m}; \quad (7.236)$$

$$x_j \geq 0, \quad (7.237)$$

где z_{ij} можно рассматривать как часть ресурса i , направленного для производства j .

Поскольку решение задачи (7.235) — (7.237) зависит от ω и

$$\bar{Z} = \{z_{ij}, i = \overline{1, m}; j = \overline{1, n}\},$$

т. е. $\bar{X}$ является функцией от $\bar{Z}$ и ω , то можно рассмотреть задачу: минимизировать

$$F(\bar{Z}) = E \sum_{j=1}^n c_j(\omega) x_j(\bar{Z}, \omega) \quad (7.238)$$

при ограничениях

$$\sum_{j=1}^n z_{ij} = 1, \quad i = \overline{1, m}; \quad (7.239)$$

$$0 \leq z_{ij} \leq 1, \quad i = \overline{1, m}; \quad j = \overline{1, n}. \quad (7.240)$$

Решая задачу (7.238) — (7.240), получим оптимальное решение

$$\bar{Z} = \{z_{ij}, i = \overline{1, m}; j = \overline{1, n}\}.$$

В некоторых случаях для задачи (7.235) — (7.237) можно определить функцию распределения величины $\sum_{j=1}^n c_j(\omega) x_j(\bar{Z}, \omega)$, определить математическое ожидание $F(\bar{Z}) = E \sum_{j=1}^n c_j(\omega) x_j(\bar{Z}, \omega)$ и решить задачу нелинейного программирования.

3. Пусть известно математическое ожидание $\omega = E\omega$. Тогда задачу (7.211) — (7.213) можно заменить задачей нелинейного программирования: минимизировать

$$f_0(\bar{X}, E\omega) \quad (7.241)$$

при ограничениях

$$f_i(\bar{X}, E\omega) \leq 0, \quad i = \overline{1, m}; \quad (7.242)$$

$$\bar{X} \in X. \quad (7.243)$$

Если $f_i(\bar{X}, \omega)$, $i = \overline{1, m}$, — линейные функции относительно ω , то решение задачи (7.241) — (7.243) совпадает с решением задачи (7.211) — (7.213), а при малой дисперсии ω ее решение будет близким к решению задачи (7.211) — (7.213).

Если выполняется условие

$$F_i(\bar{X}) = E f_i(\bar{X}, \omega) \leq f_i(\bar{X}, E\omega), \quad i = 0, 1, \dots, m, \quad (7.244)$$

то любое допустимое решение задачи (7.241) — (7.243) является допустимым для задачи (7.211) — (7.213) и

$$\min F_0(\bar{X}) \leq \min f_0(\bar{X}, E\omega). \quad (7.245)$$

В других случаях решение задачи (7.241) — (7.243) существенно отличается от решения задачи (7.211) — (7.213).

Для решения задач оптимизации с негладкими выпуклыми функциями цели и ограничений в условиях неточной информации об этих функциях и их производных применяется метод проектирования стохастических квазигradientов. *Стохастический квазигradient* — это случайный вектор, математическое ожидание которого в некотором смысле близко к gradientу или обобщенному gradientу функции, т. е. это вектор, являющийся их статистической оценкой. В стохастических квазигradientных методах за направление спуска (направление движения к экстремуму) выбираются некоторые случайные векторы.

Пусть дана задача:
минимизировать

$$f(x_1, \dots, x_n) \quad (7.246)$$

при ограничениях

$$\bar{X} = (x_1, \dots, x_n) \in X, \quad (7.247)$$

где $f(\bar{X})$ — выпуклая вниз функция (точные значения $f(\bar{X})$ и ее производных вычислить нельзя); X — выпуклое замкнутое множество.

Пусть (Ω, A, P) — некоторое исходное вероятностное пространство, ω — множество элементарных событий, а $\bar{X}^l(\omega)$, $l=0, 1, \dots$, — последовательность случайных точек, определенная по правилу

$$\bar{X}^{l+1} = \pi_x(\bar{X}^l - \rho_l \gamma_l \xi^l(\omega)), \quad l = 0, 1, \dots, \quad (7.248)$$

где π_x — оператор проектирования на множество X ; $\bar{X}^0$ — производная точка пространства R^n ; ρ_l — шаг движения; γ_l — нормирующий множитель (величины ρ_l и γ_l измеримы относительно σ -подалгебры G_l); $\xi^l(\omega)$ — случайный вектор, условное математическое ожидание которого

$$E(\xi^l(\omega) | \bar{X}^0, \dots, \bar{X}^l) = a_l \hat{f}_x(\bar{X}^l) + \bar{B}^l, \quad l = 0, 1, \dots; \quad (7.249)$$

a_l и $\bar{B}^l = (b_1^l, \dots, b_n^l)$ — неотрицательная случайная величина и случайный вектор, зависящие от последовательности $(\bar{X}^0, \dots, \bar{X}^l)$ или, более точно, измеримые относительно σ -подалгебры G_l , индуцированной семейством случайных величин $(\bar{X}^0, \dots, \bar{X}^l)$; $\hat{f}_x(\bar{X}^l)$ — обобщенный gradient, т. е. вектор, удовлетворяющий для любых $\bar{Z}$ неравенству

$$f(\bar{Z}) - f(\bar{X}^l) \geq (\hat{f}_x(\bar{X}^l), \bar{Z} - \bar{X}^l).$$

Вектор ξ^l , удовлетворяющий условию (7.249) с точностью до a_l и $\bar{B}^l$, в среднем совпадает с вектором обобщенного gradientа. Если $a_l \equiv 1$, $\bar{B}^l \equiv 0$, то вектор ξ^l является стохастическим обобщенным gradientом (стохастическим квазигradientом), а процедура (7.248) называется методом проектирования стохастических квазигradientов.

Процедура (7.246) — (7.249) может сходиться к решению задачи (7.246) — (7.247) в различном смысле: по вероятности, в сред-

нем квадратичном, с вероятностью 1. В последнем случае за приближенное решение задачи при большом l можно принять $\bar{X}^l$.

Процедура (7.248), (7.249) используется при построении прямых методов решения задач стохастического программирования.

Стохастические задачи на безусловный экстремум позволяет решать метод стохастической аппроксимации. В этом методе при минимизации целевой функции $Ef(\bar{X}, \omega)$ направлением спуска выбирают антиградиент функции $f(\bar{X}, \omega)$ вместо неизвестного антиградиента $-F_x(\bar{X})$, где $F(\bar{X}) = Ef(\bar{X}, \omega)$, т. е. вместо обычного градиентного метода рассматриваются итеративные процедуры поиска, определяемые соотношениями

$$\bar{X}^{l+1} = \bar{X}^l - \rho_l f_x(\bar{X}^l, \omega), \quad l = 0, 1, \dots$$

Если при каждом ω градиент $f_x(\bar{X}^l, \omega)$ вычислить нельзя, то градиент определяется численно и

$$\bar{X}^{l+1} = \bar{X}^l - \rho_l \sum_{j=1}^n \frac{f(\bar{X}^l + \Delta_l e^j, \omega^{l_0}) - f(\bar{X}^l, \omega^{l_0})}{\Delta_l} e^j, \quad l = 0, 1, \dots, \quad (7.250)$$

где e^j — орт j -й оси; ω^{l_s} , $s = 0, 1, \dots, n$ — результаты независимых по $l = 0, 1, \dots$ наблюдений случайной величины ω ; ρ_l — длина шага в l -й итерации; Δ_l — смещение по осям. Можно считать $\omega^{l_0} = \omega^{l_1} = \dots = \omega^{l_n} = \omega^l$.

Метод (7.250) является частным случаем процедуры (7.248) — (7.249) при

$$\xi^l = \sum_{j=1}^n \frac{f(\bar{X}^l + \Delta_l e^j, \omega^{l_j}) - f(\bar{X}^l, \omega^{l_0})}{\Delta_l} e^j.$$

Глава 8

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

8.1. ОБЩАЯ ХАРАКТЕРИСТИКА СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

На современном этапе развития единого интегрированного процесса «наука — техника — производство» все более важная роль отводится системам автоматизированного проектирования (САПР) — одному из важнейших факторов ускорения научно-технического прогресса и эффективного использования ЭВМ в области управления сложными системами.

Понятие САПР включает в себя множество различных направлений, основное из которых связано с автоматизацией проектирования изделий, технических устройств и систем.

Проектирование сложных объектов — весьма трудоемкий процесс, в котором наряду со значительной долей творческого труда имеет место большой объем операций, именуемых рутинными. Проектирование в целом можно определить как процесс составления описания, необходимого для создания в заданных условиях еще

не существующего объекта на основе его первичного описания. Проектирование на современном этапе, как правило, связано с оптимизацией заданных характеристик объекта и алгоритма его функционирования. Это означает, что из множества возможных вариантов объекта, системы и т. п. необходимо выбрать оптимальный в заданном смысле вариант. При этом, как правило, задачи оптимизации проектно-конструкторских решений носят многокритериальный характер, вследствие чего приходится отыскивать компромиссные варианты путем трудоемкого и сложного анализа множества экономических, технических, технологических и других показателей критериев качества. Сами же критерии качества обычно разносторонни, противоречивы и не могут быть формализованы в виде некоторой скалярной функции. Вместе с тем аналитическая теория решения задач оптимизации с векторными критериями оптимальности многокритериальных задач в настоящее время не обладает достаточно удобными методами. Вследствие этого оптимизация проектно-конструкторских решений обычно осуществляется по нескольким независимым критериям с широким использованием методов экспертных оценок. Важно подчеркнуть, что глобальный экстремум основного показателя качества должен быть достигнут за счет взаимосвязанной оптимизации всего проектируемого комплекса в целом, а не за счет оптимизации отдельных узлов.

Отмеченные особенности современного проектирования делают этот процесс чрезвычайно сложным с точки зрения получения конечных оптимальных результатов. К этому следует добавить огромные масштабы и трудоемкость проектно-конструкторских работ. Например, в области строительства в нашей стране функционирует свыше 1700 проектно-изыскательных организаций, в которых занято более 750 тыс. человек. Ежегодно в отрасли строительства выполняется огромный объем проектных работ на сумму около трех миллиардов рублей. Традиционная технология проектирования, основанная на безмашинной (ручной) форме организации проектирования, использующая простейшие технические средства проектирования — кульманы, логарифмические линейки, готовальни и т. д., не может удовлетворять требованиям ускорения научно-технического прогресса.

Существенное усложнение проектируемых объектов, многовариантность проектирования, противоречивость критериев оптимальности, дефицит трудовых ресурсов, огромные масштабы и краткие сроки проектно-конструкторских работ остро поставили вопрос широкого применения автоматизированных технологий проектирования объектов, изделий, систем во всех областях народного хозяйства. КПСС и Советское правительство придают огромное значение развитию методов и средств автоматизированного проектирования. В Основных направлениях экономического и социального развития СССР на 1986—1990 годы и на период до 2000 года, принятых на XXVII съезде КПСС, говорится:

«Развивать специализацию проектных и изыскательских организаций... Шире внедрять автоматизированную систему проекти-

рования». Автоматизация проектирования призвана внести существенный вклад в решение задач, сформулированных в новой редакции Программы Коммунистической партии Советского Союза: «...в экономической области — подъем народного хозяйства на принципиально новый научно-технический и организационно-экономический уровень, перевод его на рельсы интенсивного развития; достижение высшего мирового уровня производительности общественного труда, качества продукции и эффективности производства...».

В области строительства вопросы широкого внедрения автоматизации проектирования стоят очень остро. При сохранении традиционных методов проектирования с использованием реальной численности проектировщиков в самое ближайшее время можно удовлетворить не более 50 % потребностей прироста объемов проектно-изыскательных работ, вызванного развитием капитального строительства. Без использования методов и средств автоматизированного проектирования невозможно удовлетворить возрастающие требования к качеству проектов, снижению стоимости и материалоемкости проектируемых объектов, повышению темпов проектных работ. Перед проектными организациями поставлена задача довести уровень автоматизации проектных работ в целом по стране до 20 %.

Основными целями внедрения САПР в отрасли строительства являются: существенное повышение производительности труда при выполнении проектных работ; снижение капитальных затрат на строительство; сокращение расходов материалов и технико-экономических ресурсов.

Под автоматизированным проектированием понимается проектирование, при котором преобразования описаний объекта и алгоритма его функционирования, а также представление описаний на различных языках и в различной форме осуществляются взаимодействием человека и ЭВМ. Системы, основанные на использовании ЭВМ и экономико-математических методов, предусмотренные для автоматизации проектно-конструкторских работ, получили название систем автоматизированного проектирования (САПР). САПР — организационно-техническая система или иначе — система «человек — машина», в которой сбор, хранение и обработка информации, необходимой для эффективного и качественного проектирования, осуществляются на основе комплекса средств автоматизации проектирования, взаимосвязанных с необходимыми подразделениями проектных организаций или коллективом специалистов — пользователями системы.

Основу комплекса технических средств (КТС) САПР составляет ЭВМ и связанное с ней разнообразное периферийное оборудование. Характерной особенностью САПР является комплексное использование ЭВМ как для автоматизации творческого труда, так и для автоматизации рутинных операций. В первом направлении ЭВМ используется как мощная логико-вычислительная система для автоматизированного поиска и выбора оптимального варианта проектно-конструкторских решений на базе проектирова-

ния. Во втором направлении ЭВМ используется как эффективное средство управления разнообразными сложными техническими устройствами, обеспечивающими автоматическое изготовление чертежей, смет, вариантов схем размещения элементов объекта проектирования, оборудования и т. п. ЭВМ и САПР позволяет широко использовать численные методы исследования и оптимизации систем и объектов проектирования, приближенные аналитические и численные методы построения субоптимальных проектных решений с целью допустимого упрощения задач и экономии машинного времени. В связи с изложенным большое значение приобретают правильное разделение функций человека и ЭВМ в САПР, выбор эффективных средств взаимодействия проектировщиков с ЭВМ.

Задача заключается в том, чтобы наиболее полно использовались возможности человека и ЭВМ, дополняющие друг друга в сложном процессе проектирования. К кругу вопросов, решение которых недоступно машине и, следовательно, возлагается на человека, относятся, прежде всего, формулировка конечной цели, декомпозиция ее на отдельные иерархические уровни. В отличие от ЭВМ человек способен оценить те аспекты решаемых задач, которые трудно поддаются количественной оценке, например социальные, эстетические и другие факторы. Человек способен формулировать критерии оценки решаемых задач, учитывать факторы, возникающие в процессе проектирования, включать их в механизм оценки, использовать опыт, интуицию, воображение, принимать решение при неполных исходных данных. Короче говоря, главная отличительная черта человека, определяющая его роль и место в САПР как человеко-машинной системе,— способность к неформальному мышлению, к изменению стратегии действий в любой момент процесса проектирования, гибкость мышления.

В свою очередь, ЭВМ, в отличие от человека, способна хранить огромные объемы информации при очень коротком времени доступа к ней, имеет чрезвычайно высокое быстродействие, высокие надежность и производительность в выполнении шаблонных операций. ЭВМ в САПР реализует сложнейшие законы управления многими техническими, в частности, исполнительными устройствами.

Следует, однако, подчеркнуть, что чрезвычайная сложность процесса проектирования не позволяет однозначно разделить его на операции, выполняемые ЭВМ, и операции, выполняемые человеком. Обобщенно можно заключить, что ЭВМ способна выполнять те операции процесса проектирования, которые могут быть формализованы, т. е. строго описаны на алгоритмическом языке. Все, что не поддается формализации, выполняется человеком.

Практика разработки и внедрения САПР выработала эффективную форму взаимодействия человека с ЭВМ в процессе проектирования. Такой формой стал диалоговый режим, позволяющий проектировщику активно вмешиваться в работу ЭВМ на любой стадии ее работы, совершенствуя процесс проектирования на основе неформализуемых факторов. Диалоговый режим — наиболее

удобная форма сочетания огромных вычислительно-логических возможностей ЭВМ и неформального, творческого мышления человека.

Диалоговый режим работы САПР потребовал разработки сложного комплекса технических устройств, которая еще далека от завершения.

Развитие САПР в целом связано с необходимостью разработки специализированных ЭВМ, специальных периферийных устройств, специального математического обеспечения для автоматизации проектирования, оптимизации проектно-конструкторских решений. Показатели, по которым осуществляется оптимизация проектных решений на той или иной стадии проектирования, выбираются с учетом назначения проектируемой системы. Например, при проектировании динамических систем управления оптимизация проектных решений на всех стадиях проектирования осуществляется по показателям: экономическим (стоимость); динамическим (точность, быстроедействие, показатели переходных процессов при типовых возмущениях); физической реализации (масса, габариты, надежность, ремонтпригодность и т. п.).

Важно подчеркнуть, что оптимизация проектных решений в САПР должна осуществляться на всех стадиях проектирования, предусмотренных ГОСТом на Единую систему конструкторской документации: проектного предложения, эскизного проекта, технического проекта, рабочего проекта.

Являясь прогрессивным и эффективным направлением использования ЭВМ для автоматизации деятельности человека, САПР широко внедряются в различные отрасли народного хозяйства: строительства, машино-, судо-, приборостроение и т. п. Эти системы выполняют проектирование частей и узлов сложных изделий и систем управления, оптимизируя проектно-конструкторские решения по определенному кругу показателей. Важнейшая задача на современном этапе научно-технического прогресса заключается в разработке и широком внедрении таких САПР, которые должны полностью охватывать проектирование любых сложных комплексов технических устройств и систем в целом — от выдачи задания на проектирование до автоматического изготовления чертежей, узлов и деталей объектов проектирования. Важно подчеркнуть, что результаты проектирования в САПР могут представляться в виде информации на машинных носителях, используемых для управления технологическим оборудованием при изготовлении проектируемых изделий. Это открывает путь к созданию полностью автоматизированных производств и переходу к безлюдной технологии. Подробно этот вопрос рассмотрен в гл. 9.

8.2. СХЕМА ПРОЦЕССА ПРОЕКТИРОВАНИЯ И ЭЛЕМЕНТНЫЙ СОСТАВ САПР

Прежде чем перейти к рассмотрению структуры и состава САПР, остановимся на кратком описании процесса автоматизированного проектирования.

Исходная информация для проектирования как первичное описание объекта проектирования в САПР хранится на машинных носителях — дисках, лентах, гибких дисках. Там же хранятся математические модели проектируемых объектов или их деталей, узлов, частей. Используя средства отображения на основе электронно-лучевой трубки — дисплей, проектировщик имеет прямой и удобный доступ к этой информации. На основе первичного описания объекта проектировщик конструирует его окончательное описание в форме, необходимой для создания объекта. При этом первичное описание подвергается неоднократному преобразованию с последовательной оптимизацией характеристик объекта на базе математических моделей, методов и алгоритмов процесса проектирования, входящих в систему автоматизированного проектирования. Технические средства САПР позволяют контролировать промежуточные результаты проектирования, последовательно улучшая их с целью удовлетворения критериям качества. Для получения желаемого варианта проектно-конструкторского решения аппаратно-программным путем необходимо создать требуемое количество документов: чертежей, схем, графиков, информации на машинных носителях. В результате весь процесс автоматизированного проектирования на базе ЭВМ и экономико-математических методов разбивается на множество проектных процедур $P\{p_1, p_2, \dots, p_N\}$. Каждая проектная процедура осуществляется на базе взаимодействия трех множеств: множества математических моделей $M\{M_1, M_2, \dots, M_n\}$; множества операций над моделями $O\{O_1, O_2, \dots, O_n\}$; множества критериев проектирования $K\{K_1, K_2, \dots, K_n\}$.

Каждая проектная процедура $P_i(O_i, K_i, M_i)$ переводит модель проектируемого объекта в состояние $\forall M_i \in M \exists P: M_i \rightarrow M_{i+1}$, где $M_i \subset M_{i+1}$, $K_i \subset K_{i+1}$, $O_i \subset O$, $i \in [0, n]$, $l \in [0, L]$.

Взаимодействия M, O, K осуществляются как итерационный процесс (рис. 8.1), состоящий из циклов. Циклы повторяются на каждом этапе проектирования.

В состав цикла входят: 1) генерация типовых проектных решений (ТПР)_j; 2) предварительный анализ (ТПР)_j для новых условий, заданных функционалом качества I_{0j} и техническим заданием (ТЗ)_{i-1}, составленным по результатам предыдущей процедуры, с проверкой выполнения функционального неравенства $I_{ji} \leq I_{0i}$; 3) оптимизация параметров по критериям, заданным в (ТЗ)_{i-1}; 4) моделирование проектируемого объекта или его элементов.

Если имеющиеся ТПР не удовлетворяют требованиям проектируемого

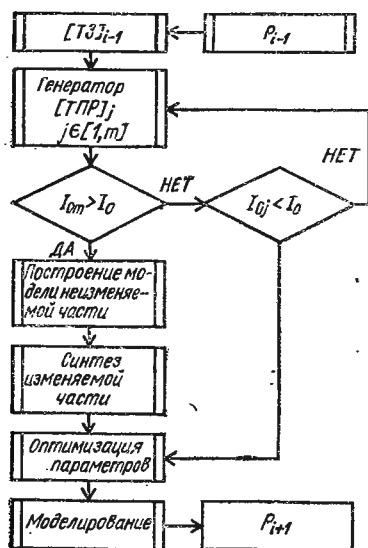


Рис. 8.1. Схема алгоритма одного цикла проектирования в САПР

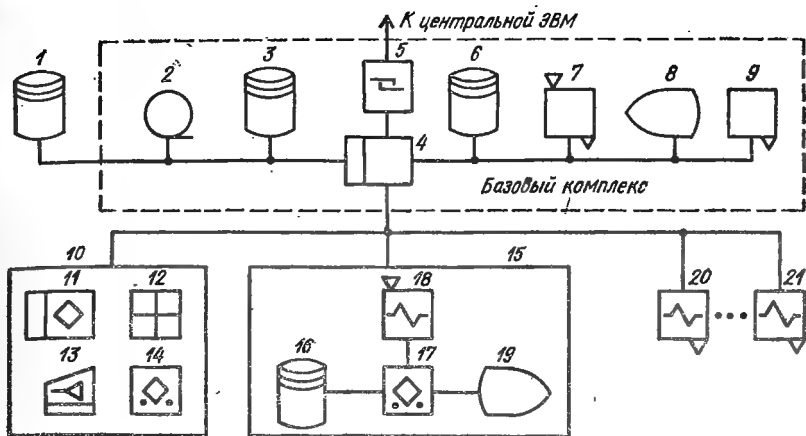


Рис. 8.2. Система автоматизированного проектирования:

1, 3 — устройства внешней памяти на магнитных дисках; 2 — то же, на магнитной ленте; 4, 5 — устройство связи вычислительных машин; 7 — устройство ввода-вывода на перфокарте; 8 — алфавитно-цифровой видеотерминал; 9 — устройство печати; 10 — графический дисплей (11 — специпроцессор; 12 — постоянное запоминающее устройство; 13 — электронно-лучевая трубка; 14 — электронно-клавишная вычислительная машина); 15 — устройство кодирования и ввода чертежей и эскизов (16 — устройство внешней памяти на магнитных дисках; 17 — микроЭВМ; 18 — устройство ввода графической информации; 19 — дисплей); 20, 21 — графопостроители

объекта, в САПР предусматривается формирование математической модели неизменяемой части объекта, синтез структуры и параметров нетиповой части проектируемого объекта и его элементов.

Описанный алгоритм цикла проектирования реализуется в САПР с помощью комплекса технических средств с использованием ЭВМ и экономико-математических методов. Один из вариантов состава технических средств САПР приведен на рис. 8.2. Постоянное запоминающее устройство (ПЗУ) в отличие от оперативного (ОЗУ) содержит постоянную информацию, которая не может быть заменена другой в процессе эксплуатации.

Система позволяет: формировать задание на проектирование; выводить данные на графопостроитель, графический дисплей, алфавитно-цифровые устройства; вводить информацию в устройства памяти с устройств кодирования, дисплеев, центрального вычислителя; решать широкий круг задач с приемом, хранением и переработкой информации; вести диалоговый режим проектирования. Система обеспечивает независимую работу нескольких пользователей со многих пультов при решении нескольких задач. Проектировщики работают за устройствами кодирования и графическими дисплеями. Чертежи и эскизы в память машины вводятся с помощью устройств кодирования, содержащих устройства считывания и преобразования графической информации. Описания привязочных точек символов или графических групп хранятся в машине. Оператор «скальвает» эти привязочные элементы и с помощью клавиатуры и матрицы символов («меню») задает признаковую информацию.

Устройство ввода чертежей и эскизов (15) включает: планшет в виде матрицы, размещенной на специальном основании; съемник; пульт управления; алфавитно-цифровой дисплей. Съемник координат имеет гибкую связь и может перемещаться по поверхности матрицы. Ввод режимов работы и задание символической информации осуществляются с пульта управления, свободно перемещающегося по поверхности координатной матрицы. Чертежи больших размеров, превышающих габариты планшета, могут вводиться по частям с последующим «сшиванием», осуществляемым программно.

Изображения на чертежах и эскизах могут быть привязанными к координатной сетке, что упрощает их, так как в этом случае не требуется проставлять размеры, снижаются требования к точности выполнения эскизов, вводимых с кодировщика.

Описанная система проектирования может работать как в непосредственной связи с центральной ЭВМ, так и в автономном режиме, в котором ввод-вывод информации осуществляется с использованием накопителя на гибких магнитных дисках.

МикроЭВМ (17) в составе устройства кодирования позволяет существенно разгрузить ЭВМ базового комплекса (4). МикроЭВМ решает следующие задачи: преобразование значений координат из системы координат планшета в систему координат чертежа; округление значений координат в заданный шаг координатной сетки; преобразование информации для передачи ее в ЭВМ; контроль и выдачу диагностических сообщений, управление работой накопителя на гибких магнитных дисках и работой алфавитно-цифрового дисплея.

8.3. СТРУКТУРА СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ. СТАДИИ СОЗДАНИЯ САПР

САПР относится к категории сложных систем, синтез структуры которых осуществляется на базе системного анализа. Под *системным анализом* понимается совокупность методологических средств, используемых для подготовки и обоснования решений по сложным проблемам в различных областях человеческой деятельности, в том числе в области разработки сложных организационно-технических систем. Не лишне подчеркнуть, что системный анализ не может быть безоговорочно отнесен к строго научным методам, так как в нем широко используются неформализуемые методы и приемы, интуиция, личный опыт исследователя, экспертные оценки и суждения. К такому неформальному аппарату приходится прибегать потому, что синтез сложной системы обычно проводится в условиях недостаточной исходной информации, выдвижения альтернативных вариантов решения поставленной проблемы. Вместе с тем системный анализ широко использует прикладные математические дисциплины, особенно методы математического программирования, рассмотренные в предыдущей главе. Эти математические дисциплины составляют теоретическую базу

системного анализа, а его техническую базу составляют электронно-вычислительные машины и информационные системы.

Системный анализ включает в себя следующие основные этапы:

- 1) постановка задачи, выявление и формулирование конечных целей;
- 2) первичная структуризация системы с целью определения ее границ;
- 3) составление математической модели системы на базе математического описания ее объектов и процессов, протекающих в них;
- 4) исследование математической модели с целью выявления соответствия системы заданным целям и критериям.

При синтезе САПР как сложной системы необходимо определить целесообразное число уровней ее иерархической структуры, установить связи между элементами каждого уровня и между уровнями, разработать схемы управления и схемы принятия решений, упорядочить информационные потоки в системе, исходя из ее конечных целей, определить требуемый комплекс технических средств.

Как всякая сложная система, САПР представляется как взаимосвязанная совокупность тесно взаимодействующих подсистем. При этом *подсистемой* называется часть САПР, выделенная по функциональному, структурному или иному признаку, имеющая конкретные цели и задачи, достижение которых необходимо для эффективного функционирования всей сложной системы как единого целого.

В соответствии с изложенными предпосылками в настоящее время определен типовой состав САПР как совокупность подсистем, каждая из которых состоит из компонентов САПР, объединенных общей для данной подсистемы целевой функцией и обеспечивающих ее функционирование.

По назначению подсистемы САПР подразделяются на проектирующие и обслуживающие.

Проектирующие подсистемы выполняют проектные процедуры и операции, такие как проектирование деталей и узлов, частей зданий, технологическое проектирование. Название проектирующих подсистем соответствует их назначению, например «Подсистема технологического проектирования».

Обслуживающие подсистемы обеспечивают работоспособность проектирующих систем. Примерами обслуживающих подсистем являются «Подсистема информационного поиска», «Подсистема графического отображения объектов проектирования».

По характеру связи проектных процедур и операций с объектом проектирования проектирующие подсистемы подразделяются на объектно-ориентированные (объектные) и объектно-независимые (инвариантные). *Объектными* являются подсистемы, которые выполняют одну или несколько проектных процедур или операций, зависящих от объекта проектирования. *Инвариантные* подсистемы выполняют унифицированные проектные процедуры и операции, непосредственно не зависящие от конкретного объекта проектирования.

Каждая подсистема включает в себя компоненты САПР — элементы, обеспечивающие функционирование подсистемы и выполняющие определенные функции в подсистеме. Все компоненты подсистемы объединены общей целевой функцией данной подсистемы.

Компонентами подсистемы САПР являются:

1) методическое обеспечение — документы, отражающие состав, правила выбора и эксплуатации средств автоматизации проектирования;

2) лингвистическое обеспечение — языки проектирования, терминология;

3) математическое обеспечение — совокупность математических моделей, методов и алгоритмов, используемых при решении задач в подсистеме;

4) программное обеспечение — совокупность программ на машинных носителях, документов с текстами программ, эксплуатационных документов;

5) информационное обеспечение — совокупность документов и их записей на машинных носителях, описывающих стандартные проектные процедуры, типовые проектные решения, типовые элементы, комплектующие изделия, материалы и другие данные;

6) техническое обеспечение — комплекс технических средств, включающий средства вычислительной, организационной техники, средства передачи информации, измерительные и другие устройства и средства;

7) организационное обеспечение — совокупность положений, приказов, квалификационных характеристик, штатных расписаний и других документов, определяющих организационную структуру подразделений САПР и их взаимодействие со средствами автоматизации проектирования.

Каждая подсистема является единым структурным целым, что обеспечивается взаимными связями компонентов, входящих в подсистемы. САПР в целом также является единой структурой, что обеспечивается взаимными связями компонентов всех подсистем, образующих систему автоматизированного проектирования.

САПР как сложная система, ее подсистемы и компоненты создаются на основе основных системных принципов: системного единства; развития; совместимости; стандартизации. *Системное единство* заключается в том, что взаимосвязи между подсистемами при создании, функционировании и развитии САПР должны обеспечивать целостность всей системы. *Принцип развития* подразумевает, что создание и функционирование САПР должно осуществляться с учетом пополнения, совершенствования и обновления подсистем и компонентов САПР. *Принцип совместимости* требует, чтобы языки, коды, символы, информационные и технические характеристики структурных связей между подсистемами, средствами обеспечения и компонентами САПР обеспечивали совместное функционирование всех подсистем, сохраняя открытую структуру в целом. Наконец, *принцип стандартизации* устанавливает требования унификации, типизации и стандартизации подсистем и компонентов, инвариантных к проектируемым объектам и отраслевой

специфике. Принцип стандартизации также устанавливает правила упорядочения работ по созданию и развитию САПР.

Процесс создания САПР и ее подсистем как сложных систем состоит из отдельных стадий, увязанных друг с другом.

Содержание отдельных стадий, характер работ по стадиям и содержание документации на отдельных стадиях определяют ГОСТ 23501.0—79—23501.603—84.

В настоящее время установлены следующие стадии создания САПР: 1) предпроектные исследования; 2) техническое задание; 3) техническое предложение; 4) эскизный проект; 5) технический проект; 6) рабочий проект; 7) изготовление, отладка и испытание; 8) ввод в действие.

Предпроектное исследование заключается в обследовании проектной организации, оформлении, согласовании и утверждении технического отчета. На стадии технического задания разрабатывается, согласовывается и утверждается техническое задание на разработку САПР. Техническое предложение включает в себя выбор и обоснование рационального варианта САПР, разработку соответствующего комплекта документации, его согласование и утверждение. На стадии эскизного проекта разрабатываются принципиальные решения по созданию САПР, а на стадии технического проекта разрабатываются окончательные решения. Разработки рабочей документации по САПР осуществляются на стадии рабочего проекта. Эскизный, технический и рабочий проекты согласовываются и утверждаются в установленном порядке. На стадии изготовления, отладки и испытания осуществляются создание и отладка компонентов САПР, монтаж, наладка и испытание комплекса технических средств автоматизации проектирования, организуется подготовка к вводу системы в действие. На стадии ввода в действие выполняются приемочные испытания САПР и опытная эксплуатация.

Подсистемы и компоненты САПР реализуются на стадиях № 1 — № 6. Для отдельных САПР стадии технического предложения и эскизного проекта не обязательны, а предусмотренные этими стадиями работы выполняются в техническом проекте. Для подсистем САПР стадии № 1, 3, 4 и 5 не являются обязательными, а работы, предусмотренные на этих стадиях, выполняются при рабочем проектировании.

При создании САПР необходимо шире использовать типовые подсистемы, разработанные ранее и ориентированные на проектирование объектов, их составляющих и технических процессов одного класса (вида). В таком случае основная работа при создании САПР сводится к привязке типовых подсистем к конкретным условиям разрабатываемой САПР на стадии рабочего проекта. При этом стадии предпроектных исследований эскизного и технического проектов не обязательны.

8.4. ОСНОВНЫЕ КОМПОНЕНТЫ СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Компоненты САПР, перечисленные в предыдущем параграфе, тесно взаимодействуют друг с другом, обеспечивают функционирование отдельных подсистем и всей системы в целом. Остановимся кратко на содержании основных компонентов САПР.

Лингвистическое обеспечение САПР (ЛО САПР) — совокупность языков, включающая термины и определения, правила формализации естественного языка, методы преобразования текстов, используемых при разработке и эксплуатации САПР в процессе обмена информацией между человеком и ЭВМ. ЛО САПР разделяется на языки программирования, проектирования и управления. *Языки программирования* — алгоритмические языки, изучаемые в курсе «Программирование». *Языки проектирования* — проблемно-ориентированные языки, используемые для обмена информацией об объектах проектирования между проектировщиком и ЭВМ. *Языки управления* — языки, используемые для управления технологическим оборудованием и периферийными устройствами ЭВМ.

В качестве языков программирования в САПР используются алгоритмические языки высокого уровня — ПЛ/1, ФОРТРАН, АЛГОЛ-60; машинно-ориентированные языки — языки АССЕМБЛЕР. Весьма перспективны для использования в САПР языки ПАСКАЛЬ и АДА.

Выбор конкретного языка программирования определяется свойствами языка и условиями разработки и эксплуатации САПР. Для программ с преобладанием логических операций над адресами и другими двоичными кодами целесообразнее использовать языки АССЕМБЛЕР, ПАСКАЛЬ, АДА, ПЛ/1. Языки АССЕМБЛЕР более предпочтительны при необходимости жесткой экономии машинного времени при эксплуатации САПР, а алгоритмические языки высокого уровня предпочтительнее при необходимости сокращения сроков разработки программ. Для разработки прикладных программ САПР наиболее широко используется язык ФОРТРАН.

Языки проектирования подразделяются на входные, выходные, языки сопровождения и внутренние языки.

Входные языки используются для представления сведений об объектах и целях проектирования. Сведения об объектах представляются на языках описания объекта (ЯОО) и языках описания заданий (ЯОЗ), являющихся составными частями входных языков.

Средствами ЯОО описывается структура и исходные значения параметров. При этом в зависимости от формы, в которой даются сведения об объекте, используются различные типы ЯОО: графический язык — для представления сведений об объекте, заданных в виде чертежей; схемный (табличный) язык — для сведений, заданных в форме списков элементов структуры; язык моделирования — если исходные данные представляют собой описание действий, реализуемых в объекте.

Средствами ЯОЗ передаются задания для ЭВМ в виде сведений о том, какие задачи и при каких условиях должны решаться. Этот

язык по возможностям и грамматике близок к языкам управления заданиями в операционных системах ЭВМ, отличаясь проблемной ориентацией.

Результаты реализации программ САПР представляются с помощью *выходных языков*, обеспечивающих формирование таблиц, графиков, чертежей, диаграмм, текстовых сообщений, диагностических сообщений об ошибках.

Непосредственное обращение пользователя к ЭВМ в процессе решения задач осуществляется на *языках сопровождения*, содержащих специфические средства редактирования и управления. Язык сопровождения может содержать также элементы входного и выходного языков. Языки сопровождения обеспечивают работу пользователя как в диалоговом, так и в пакетном режимах работы. По этому признаку языки сопровождения подразделяются на диалоговые и недиалоговые.

Строгие правила для языковых конструкций необходимы не только для общения человека с ЭВМ, но и на некоторых этапах обработки информации внутри ЭВМ. С этой целью в лингвистическом обеспечении имеется унифицированный для конкретной САПР языковой интерфейс — единый *внутренний язык*. С помощью этого языка, например, удовлетворяются требования взаимной независимости инвариантной и объектно-зависимой частей программного обеспечения САПР.

Математическое обеспечение САПР (МО САПР). В состав математического обеспечения входят математические модели объектов проектирования, методы и алгоритмы выполнения проектных операций и процедур. Математическое обеспечение (МО) базируется на математическом аппарате моделирования процессов и объектов, анализе и оптимизации моделей. МО призвано также обеспечить формулировку технического задания на проектирование и синтез структуры САПР, однако эта часть МО развита слабо, а сами задачи решаются обычно в диалоговом режиме с использованием неформализуемых приемов.

Математическое обеспечение САПР подразделяется на специальную часть — специальное математическое обеспечение — и инвариантную часть. Специальное МО отражает особенности конкретных проектируемых объектов и процессов функционирования. Инвариантная часть МО предназначена для использования без тесной связи с особенностями проектируемых объектов. В состав инвариантного МО входят методы и алгоритмы многовариантного анализа и параметрической оптимизации.

Математическое обеспечение оказывает существенное влияние на качество функционирования САПР. Отметим основные требования, предъявляемые к МО САПР: универсальность; алгоритмическая надежность; точность; экономичность.

Универсальность МО характеризуется степенью его применимости к широкому классу проектируемых объектов. Свойство универсальности не имеет количественной оценки, но по этому признаку МО должно обеспечивать проектирование большинства объектов, выполняемое в конкретной САПР. Например, в области МО

САПР систем автоматического управления в настоящее время разработаны и широко применяются алгоритмы и стандартные программы расчета переходных процессов и ошибок в линейных и нелинейных системах.

Наибольшее распространение для исследования динамических систем при регулярных воздействиях нашли алгоритмы численного интегрирования Эйлера, Рунге — Кутта, Адамса, Коуэлла. Для анализа случайных процессов используются методы вычисления математических ожиданий, корреляционных функций. Численные методы принципиально способны обеспечить решение с любой заданной точностью. Однако при решении задач оптимизации нелинейных систем высоких порядков численные методы связаны с огромным объемом вычислений и в ряде случаев не обеспечивают устойчивости вычислительного процесса. В таких случаях разрабатываются и используются более экономичные и устойчивые приближенные численные и расчетно-аналитические методы, такие как матричные методы для линейных систем, метод гармонической линеаризации, различные расчетные схемы, учитывающие частные особенности проектируемых систем.

Случайные процессы в нелинейных системах исследуются с использованием метода статистических испытаний и расчетных процедур, базирующихся на каноническом разложении случайного процесса, а также методы обычной и статистической линеаризации нелинейностей. Эти методы позволяют приближенно интегрировать уравнения системы относительно фазовых координат или относительно их математических ожиданий и корреляционных моментов.

Фундаментом многих теоретических и расчетно-аналитических методов определения оптимальных управлений, как отмечалось выше, являются принцип максимума Понтрягина и метод динамического программирования Беллмана. Однако эти методы неудобны для программирования и использования в качестве стандартных процедур на цифровых электронно-вычислительных машинах. Принцип максимума, за исключением частных задач, неудобен для алгоритмизации, особенно для объектов выше второго порядка, ввиду того, что предполагает численное интегрирование сложной системы нелинейных уравнений и построение фазовых портретов в пространстве многих переменных. Метод динамического программирования в численной форме для объектов выше четвертого порядка требует больших объемов памяти ЭВМ и связан с громоздкими вычислениями.

Использование метода нелинейного программирования для оптимизации законов управления заключается в параметрической оптимизации оператора управления с заданной структурой. Целевая функция становится функцией параметров оператора управления, а дифференциальные уравнения системы входят в систему ограничений. Решение производится методом упорядоченного поиска оптимальных параметров в области, заданной ограничениями. Для каждого сочетания варьируемых параметров требуется проводить численное интегрирование дифференциальных уравнений системы или решать многомерную систему алгебраических урав-

нений, образующуюся при переходе от дифференциальных уравнений к разностным. Изложенная схема расчета также весьма трудоемка и не всегда применима.

Универсальность МО САПР существенно упрощает методику автоматизированного проектирования.

Алгоритмическая надежность МО САПР характеризует свойство математического обеспечения при его применении в заранее оговоренных условиях давать правильные результаты. Математический метод считается алгоритмически надежным, если близка к единице или равна ей вероятность получения правильных результатов с учетом оговоренных ограничений на применение метода.

Точность МО определяется степенью совпадения расчетных и истинных результатов. Алгоритмически надежные методы могут дать различную точность в пределах допустимых значений.

Количественная оценка погрешностей, вносимых различными компонентами МО, существенно затрудняется векторным характером результатов. Например, задачи оптимизации сводятся к нахождению множества координат экстремальной точки, отвечающих целевой функции. На практике существуют специальные вычислительные эксперименты, позволяющие отдельно оценить погрешности, вносимые различными компонентами математического обеспечения, алгоритмами оптимизации и анализа. Основу этих экспериментов составляют тестовые задачи. В качестве количественной оценки погрешности результата решения тестовой задачи принимаются m -норма или l -норма вектора относительных погрешностей, вычисляемые по формулам:

$$\varepsilon_{\max} = \max_{j \in \{1:m\}} |\varepsilon_j|; \quad \varepsilon_{\text{общ}} = \sqrt{\sum_{j=1}^m \varepsilon_j^2},$$

где ε_j — относительная погрешность определения j -го элемента вектора результатов; m — размерность вектора результатов.

Экономичность МО характеризуется затратами машинного времени и оперативной памяти при решении задач САПР. Объемы этих затрат часто становятся препятствием к повышению сложности проектируемых объектов в автоматизированном проектировании. Одно из эффективных средств снижения затрат машинного времени — распараллеливание вычислительного процесса. Экономия оперативной памяти достигается максимальным использованием внешней памяти.

В МО САПР важную роль играют математические методы решения задач оптимизации, рассмотренные в гл. 7. Математические модели объектов и процессов определяются областью использования САПР.

Программное обеспечение САПР (ПО САПР) представляет собой документы с текстами программ, программы на машинных носителях, эксплуатационные документы. ПО САПР подразделяется на общее программное обеспечение (ОПО) и специальное (СПО).

ОПО составляют операционные системы (ОС) ЭВМ, используемых в САПР, например ОС ЕС ЭВМ. Операционная система ЭВМ включает комплекс программ, управляющих всеми ресурсами вычислительной системы. Программы ОС используются для планирования и организации процесса обработки данных, включая ввод, вывод и управление данными, подготовку и отладку программ обслуживания вычислительной системы. Подробно функции ОС рассматриваются в дисциплинах, связанных с системным программированием.

В состав СПО входят программы выполнения проектных операций и процедур. Программы СПО выполняются в виде пакетов прикладных программ (ППП), ориентированных на решение задач отдельной подсистемы САПР. Комплекс задач подсистемы обычно решается с использованием нескольких ППП, имеющих иерархическую соподчиненность.

Каждый ППП конструируется как совокупность модулей, каждый из которых в определенной мере представляет собой самостоятельный программный компонент. Модульная структура ППП имеет большие преимущества, в частности, позволяет вести независимую разработку отдельных модулей, использовать модули для различных применений.

Программы, реализующие многократно используемые алгоритмы, образуют функциональные модули, объединенные в библиотеки. Примеры библиотек — библиотеки математических модулей элементов, численных методов решения задач, сервисных функций, отдельных проектных процедур.

Для организации взаимодействия подсистем САПР при выполнении проектных процедур, оптимизации управления процессом проектирования используется мониторингная система САПР, представляющая собой обслуживающую систему управления. В состав мониторингной системы входит основная управляющая программа-монитор, программы взаимосвязей с пользователем, диагностики ошибок и др. Мониторная система выполняет следующие основные функции: организация и управление маршрутами проектирования; организация взаимодействия с пользователями, включая диалог; обработка запросов с учетом приоритетов; защита ресурсов вычислительной системы от несанкционированного доступа. Мониторная система работает под управлением ОС ЭВМ. Обращение к ней осуществляется на процедурно-ориентированных языках описания заданий.

Рабочая программа для заданного маршрута проектирования формируется из библиотечных и оригинальных модулей. Последние отражают специфические особенности проектируемых объектов и создаются для каждого объекта с использованием языковых процессоров — программ-компиляторов, преобразующих информацию на языке описания объекта в соответствующие объектные модули рабочих программ.

В составе специального программного обеспечения САПР важную роль играет программное обеспечение машинной графики, одна из главных функций которого — обеспечение информационной

связи устройства машинной графики и прикладных программ выполнения проектных процедур. При вводе информации на входном графическом языке эта связь реализуется языковым процессором. Вывод графической информации осуществляется с использованием программ, преобразующих структуры данных прикладных программ в совокупность команд управления дисплеями и графопостроителями. В состав программного обеспечения вывода графической информации входят программы: преобразования структур данных прикладных программ в графические образы; двумерных преобразований элементов графических образов (сдвиг, поворот, масштабирование и т. п.); генерации команд управления устройствами графического вывода.

Информационное обеспечение САПР (ИО САПР) представляет собой документы, содержащие описание стандартных проектных процедур, типовых проектных решений, типовых элементов, комплектующих изделий, материалов и другие данные, а также файлы и блоки данных на машинных носителях с записью указанных документов. ИО предоставляет исходные данные, необходимые для выполнения проектных операций и процедур.

Информационное обеспечение САПР строится на основе автоматизированных банков данных (АБД), включающих базы данных (БД) и системы управления базами данных (СУБД).

Базу данных можно определить как систему данных, удовлетворяющих определенным требованиям, основными из которых являются минимальная избыточность и независимость по данным. Первое требование подразумевает коллективное многоразовое использование данных, общих для пользователей САПР, второе — независимость средств описания структур данных от прикладных программ пользователей или от изменения в структурах баз данных.

В целом БД представляют собой динамическую информационную модель объектов проектирования, состоящую из организованных специальным образом и логически связанных информационных элементов.

Система управления БД — программно-логический аппарат для организации системы хранения, занесения, обновления данных и доступа к ним.

В БД САПР хранятся две большие группы данных — архив и рабочий массив.

Архив содержит наиболее стабильную информацию — справочные данные о типах, параметрах, структуре унифицированных деталей и узлов, типовых проектах, технологических процессах. В архив включаются сведения о материалах, правилах и ограничениях, предписываемых ГОСТами и нормами, физические константы, типовые графические изображения и т. п.

Рабочий массив содержит материалы предыдущих этапов проектирования объектов, необходимые для последующих этапов, — конструкторские документы и описания на соответствующем языке. Информация, содержащаяся в рабочих массивах, обновляется значительно чаще, чем в архиве. В связи с этим архив обычно

хранится на магнитных лентах, а рабочие массивы — на магнитных дисках.

Доступ к базе данных осуществляется с использованием принятых в САПР входных языков.

Перспективы развития банков данных предусматривают превращение АБД в банки знаний, в которых наряду со справочными данными и информацией о текущих проектах будут содержаться сведения о закономерностях и методах, отражающих суть инженерных знаний.

Техническое обеспечение САПР (ТО САПР) включает в себя устройства вычислительной и организационной техники, средства передачи данных, измерительные и другие устройства или их сочетания. Основу ТО САПР составляют цифровые ЭВМ, описанные в п. 6.1. Кроме ЭВМ в САПР используется множество других сложных технических устройств. Они выбираются исходя из основного назначения технических средств САПР — организовать оперативный обмен информацией конструктора с ЭВМ. В связи с этим к ТО САПР предъявляются следующие основные требования: высокая точность вычислений; наглядность информации при взаимодействии конструктора с ЭВМ; возможность параллельной разработки отдельных частей изделия в составе общего проекта; выдача результатов проектирования на техническое оборудование; возможность использования ЭВМ конструктором непосредственно в конструкторском бюро, без обращения на вычислительный центр.

В современных САПР наиболее распространенными средствами представления результатов проектирования являются текстовые и графические дисплеи и графопостроители.

Дисплеи — устройства отображения текстовой и графической информации на экране электронно-лучевой трубки (ЭЛТ). В состав любого дисплея входит ЭЛТ, буферная память, генератор знаков (в текстовом дисплее) или генератор векторов (в графическом дисплее), аппаратные средства формирования и редактирования изображения на экране, клавиатура.

Буферная память предназначена для автономной регенерации изображения на экране, клавиатура — для ручного ввода знаков на экран, выполнения воздействий и манипуляций с изображениями на экране. В дисплейный комплекс может входить печатающее устройство.

Техническими характеристиками дисплеев являются количество знаков на экране — информационная емкость экранов, способы формирования знаков, методы выделения нужных знаков, процедуры манипуляции со знаками и изображениями, количество адресуемых точек на экране графического дисплея.

Управление режимами редактирования осуществляется с помощью клавиатуры и светового знака — курсора. При работе с ЕС ЭВМ дисплеи подключаются через мультиплексный или селекторный канал. Алфавитно-цифровые дисплеи СМ ЭВМ подключаются через контроллер — специальное буферное устройство.

Графопостроители — устройства вывода графической и текстовой информации на бумагу, используемые для выполнения черте-

жей и схем, вывода результатов вычислений в виде графиков и кривых. Графопостроители работают под управлением ЕС ЭВМ.

В состав графопостроителя входят электромеханический двухкоординатный регистрирующий построитель (ДРП) и устройство управления. ДРП включает в себя двухкоординатную систему перемещения рулонного или планшетного типа, управляемую специальными двигателями. Система обеспечивает пошаговое перемещение по бумаге пишущего узла по двум координатам в планшетном ДРП, по одной из которых происходит перемещение бумаги. Система управления может быть с буферной памятью и без нее. В буферную память принимаются графические программы, которые выбираются для обработки и управления ДРП. При отсутствии буферной памяти информация принимается непосредственно от источников ввода.

В составе графических комплексов для целей управления широко используются микроЭВМ, мини-ЭВМ и специализированные процессоры с памятью и набором периферийных устройств.

Основными техническими характеристиками графопостроителей являются: площадь рабочего поля; скорость вычерчивания; точность вычерчивания, определяемая минимальным шагом перемещения пишущего узла; толщина линий и количество цветов. Все устройства характеризуются габаритными размерами и потребляемой мощностью.

Графопостроители могут работать в автономном режиме — без связи с ЭВМ. В этом случае графические программы вводятся с перфолент или магнитных лент, подготовленных на ЭВМ.

В САПР все шире применяются *абонентские пункты* (АП) — комплекс устройств, предназначенных для работы в системе телеобработки данных на ЕС ЭВМ. Абонентские пункты представляют собой терминальную систему с программным управлением и устройствами ввода-вывода. АП устанавливаются в местах возникновения и потребления информации — в конструкторских бюро, испытательных стендах, цехах. Через модель и обычные каналы связи АП могут передавать информацию к ЕС ЭВМ, находящимся на расстоянии сотен километров от АП.

8.5. ПРИМЕРЫ СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Системы автоматизированного проектирования широко используются во многих отраслях народного хозяйства. Все более важное значение приобретает внедрение автоматизированного проектирования в строительную отрасль.

Приведем примеры САПР, используемых в строительстве.

В Научно-исследовательском институте автоматизированных систем планирования и управления в строительстве (НИИАСС) Госстроя СССР разработан вычислительный комплекс ЛИРА (линейный расчет), предназначенный для автоматизированных прочностных расчетов и выполнения элементов проектирования конструкций на базе ЕС ЭВМ: пространственно-стержневых систем,

плит, оболочек, массивных трехмерных тел и комбинированных систем — плит на упругом основании, ребристых пластин, рамно-связевых каркасов и т. п. В комплексе реализуется метод конечных элементов (МКЭ), на базе которого выполняются расчеты статических и динамических нагрузок, унификация элементов.

Для железобетонных элементов подбирается арматура по прочности, моделируется процесс их последовательного нагружения с учетом свойств железобетона. В вычислительном комплексе реализован быстродействующий алгоритм решения систем линейных уравнений, что дает возможность решать задачи с большим количеством неизвестных. Наличие развитой библиотеки конечных элементов позволяет рассчитывать объекты многих типов. Результатами расчетов на ЭВМ являются характеристики напряженно-деформированного состояния системы и сведения об армировании железобетонных элементов. Эти результаты выдаются в табличной форме, снабжаются графическими изображениями расчетных схем, эпюр усилий и схем нагрузок, представленными на графопостроителе.

Теоретическую основу вычислительного комплекса ЛИРА составляет метод конечных элементов. Библиотека конечных элементов комплекса включает программы моделирования работы конструкций различных типов на базе более 50 видов элементов. В число конечных элементов входят: стержни различных типов; треугольные, прямоугольные и четырехугольные элементы изотропной и ортотропной плит, балки-стенки, оболочки, плиты на упругом основании; прямоугольные и треугольные элементы многослойной плиты и оболочки; пространственные элементы в виде тетраэдра, параллелепипеда, восьмиугольника; одномерный, треугольный и прямоугольный осесимметричный элементы; специальные элементы, моделирующие связь конечной жесткости, упругую податливость между узлами; элементы, задающиеся численной матрицей жесткости.

Вычислительный комплекс ЛИРА функционирует на базе ЕС ЭВМ. Используются языки программирования АССЕМБЛЕР и ПЛ/1. Язык АССЕМБЛЕР используется для реализации модулей обмена с внешними устройствами, модулей распределения памяти, определения ресурсов оперативной и внешней памяти, решения системы линейных уравнений. Язык ПЛ/1 используется для реализации модулей, выполняющих ввод исходных данных, логический контроль, составление матриц жесткости конечных элементов, вычисление усилий и напряжений, динамический расчет, расчетные сочетания и армирование.

Результатами автоматизированного проектирования являются пояснительная записка, содержащая основные характеристики решаемой задачи, правила чтения результатов счета, время решения задачи. На печать также выдаются перемещения, усилия и напряжения, расчетные сочетания усилий, армирование сечений железобетонных элементов. На графопостроителе выдаются эпюры усилий для плоских рамных систем. На магнитной ленте можно получить набор данных для автоматизации таких проектных про-

цессов, как конструирование стальных конструкций и индивидуальных железобетонных элементов, проектирование зданий с выдчей чертежей монтажных схем.

Вычислительный комплекс ЛИРА повышает производительность труда проектировщиков на 30—50 %. За счет оптимизации проектируемых конструкций снижается расход бетона и арматуры на 20 %, повышается технологичность строительных конструкций.

В качестве второго примера САПР приведем технологическую линию автоматизированного проектирования производственных зданий — ТЛП ПРОМЗДАНИЙ. ТЛП представляет собой автоматизированную систему многофункционального назначения, ориентированную на комплексную автоматизацию архитектурных, конструкторских, сантехнических, электротехнических и сметных проектных работ. Система предназначена для проектирования объектов массового строительства на базе конструкций, оборудования, изделий и деталей индустриального изготовления. Функции программных систем ТЛП унифицированы и включают в себя: 1) ввод исходных данных из БАНКА ЗАДАНИЙ НА ПРОЕКТИРОВАНИЕ и БАНКА ПРОМЕЖУТОЧНЫХ ПРОЕКТНЫХ РЕЗУЛЬТАТОВ; 2) вывод результатов проектирования в БАНК ПРОМЕЖУТОЧНЫХ РЕЗУЛЬТАТОВ; 3) вывод проектных документов в табличной форме — результатов инженерных и сметных расчетов, спецификаций и ведомостей конструкций, оборудования и материалов; 4) вывод проектных документов в графической форме — схем расположения конструкций и оборудования, рабочих чертежей конструкций, общих чертежей проектируемых зданий — планов, разрезов, фасадов.

Функционально взаимодействующие программные средства ТЛП обеспечивают обмен информацией через общий для всех программ БАНК ПРОМЕЖУТОЧНЫХ РЕЗУЛЬТАТОВ, что обеспечивает значительную степень независимости программных средств от пользователей — разработчиков проектов. Каждый банк данных функционирует на базе специализированной системы управления соответствующей базой данных.

Глава 9

АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ

9.1. ОБЩАЯ ХАРАКТЕРИСТИКА И КЛАССИФИКАЦИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ

Автоматизированной системой управления (АСУ) называется организационно-техническая система, обеспечивающая эффективное функционирование объекта, в которой сбор и переработка информации, необходимой для реализации функции управления, осуществляется с активным использованием средств автоматизации и вычислительной техники. Принципы автоматизированного управления рассмотрены в гл. 1 и отражены на рис. 1.4 и 1.8. Исходя

из этих принципов видно, что АСУ — сложная система, в которой осуществляется комбинированное управление по разомкнутому циклу. Лицо, принимающее управляющие решения (ЛПР), в АСУ — это человек или группа лиц, анализирующих состояние объекта и характер влияния внешней среды с помощью ЭВМ и других технических устройств.

Объектами управления в АСУ являются обычно производственные системы, а также любые другие сложные социально-экономические системы. Назначение АСУ — обеспечить оптимальное управление объектом управления по заданным или выбранным критериям оптимальности и ограничениям. АСУ призваны поднять эффективность управления предприятиями и организациями различных отраслей народного хозяйства: промышленности, строительства, транспорта, а также эффективность управления народным хозяйством в целом. Источниками эффективности управления на базе АСУ являются сбалансированные поставки ресурсов и потребностей, оптимизация производственных процессов по различным критериям, обеспечение руководства предприятий, организаций, отраслей и т. д. своевременной и достоверной информацией о состоянии объектов управления, прогнозирование их состояний.

В зависимости от характера объектов управления АСУ подразделяются на три типа: 1) автоматизированные системы организационного управления (АСОУ); 2) автоматизированные системы управления технологическими процессами (АСУТП); 3) интегрированные автоматизированные системы управления (ИАСУ).

В АСОУ объектами управления являются производственно-экономические и социальные системы, в АСУТП — технологические процессы в сложных технических системах, в ИАСУ осуществляется комплексное управление объектами как производственно-экономического, так и технического характера.

По сложности алгоритмов обработки информации АСУ также подразделяются на три типа: 1) автоматизированные информационно-справочные системы (АИС); 2) автоматизированные системы обработки данных (АСОД); 3) автоматизированные системы, реализующие оптимальное управление (ОАСУ, или просто АСУ).

В АИС осуществляются автоматизированный сбор информации о заданных объектах, ее хранение и выдача по запросу. ЭВМ здесь используется как автоматизированный архив информации, из которого можно оперативно получить сведения об объектах и системах управления. В АИС алгоритмы управления формируют оптимальное размещение информации в памяти ЭВМ, оптимизируют процедуры ввода и вывода информации по запросу пользователей.

В АСОД осуществляются те же функции, что и в АИС, но наряду с этим осуществляется обработка информации по простейшим алгоритмам, например суммирование показателей, нахождение процентных соотношений, сопоставление результатов и т. п.

В АСУ реализуются алгоритмы оптимального управления объектом по выбранным критериям оптимизации. В этих системах имеется возможность решать сколь угодно сложные задачи управ-

ления, постановка и решение которых невозможны без использования ЭВМ и методов решения задач управления в сложных системах.

В зависимости от места объекта управления в иерархической структуре народного хозяйства различают АСУ: отдельными технологическими процессами и операциями; подразделениями предприятий (например, АСУ цеха); предприятиями (АСУП); отраслями; общегосударственные (ОГАСУ).

По характеру отдельных предприятий различают АСУ предприятий с непрерывным, дискретным (мелкосерийным и единичным) и непрерывно-дискретным производством.

Приведенная классификация АСУ условна, так как одни и те же признаки могут иметь место в системах различных типов.

По характеру функционального назначения в АСУ выделяется две части — обеспечивающая и функциональная. Обеспечивающая часть АСУ представляет собой совокупность подсистем: технического, информационного, математического, программного, лингвистического, организационного и правового обеспечения. Характеристика каждой подсистемы дается в следующем параграфе. В целом обеспечивающая часть АСУ является комплексом методов и средств, инвариантным по отношению к объекту управления, и призвана обеспечивать автоматизацию обработки информации. Автоматизация обработки данных осуществляется аналогичными методами и средствами во всех АСУ, независимо от существа задач, решаемых в системе.

Функциональная часть АСУ представляет собой комплекс административных, организационных и экономико-математических методов и средств, предназначенных для автоматизации процессов принятия решений и выработки управляющих воздействий на объект управления. Состав подсистем функциональной части определяется характером объекта управления и характером задач, решаемых в АСУ. Сложность и многокритериальность решаемых в АСУ задач требуют формировать состав подсистем функциональной части по функциям управления, обеспечивая в каждой подсистеме реализацию всех общих фаз управления — планирование, учет, контроль, анализ и регулирование. Такой подход позволяет разрабатывать алгоритмы функционирования подсистем в соответствующей фазе управления по единой методике для всех функций управления.

Для каждой функциональной подсистемы характерны свой круг решаемых задач, свой управляющий орган, цель функционирования, входы и выходы. Таким образом, каждая подсистема обладает своими свойствами и является в значительной степени автономной по информационным потокам.

Каждая функциональная подсистема характеризуется своей информационной моделью, своими экономико-математическими моделями, методами и алгоритмами решения задач управления.

Название функциональных подсистем обычно отражает круг задач управления, решаемых в подсистеме, например подсистемы: управления технической подготовкой производства; технико-эко-

номического планирования; оперативного планирования и управления; управления материально-техническим обеспечением; бухгалтерского учета и отчетности.

9.2. ОБЕСПЕЧИВАЮЩАЯ ЧАСТЬ АСУ

Техническое обеспечение АСУ (ТО АСУ) — комплекс технических средств (КТС), используемых для функционирования автоматизированной системы управления.

В соответствии со схемой обработки информации при автоматизированном управлении (см. рис. 1.8) КТС АСУ включает в себя следующие средства: сбора и регистрации информации; передачи информации; преобразования информации с целью ввода в ЭВМ; обработки информации; отображения информации с целью использования для выработки управляющих воздействий. Каждая из указанных групп имеет свой набор технических средств. Для сбора и регистрации информации широко используются регистраторы производства, для передачи — устройства и каналы связи. Преобразование информации для ввода в ЭВМ осуществляется на устройствах подготовки данных. Важнейший элемент КТС — электронно-вычислительные машины, в которых осуществляется обработка информации по заданным алгоритмам, разрабатываемым в соответствии с характером задач управления. Современные АСУ имеют широкую номенклатуру средств отображения информации, в число которых входят устройства на основе электронно-лучевой трубки (графические и текстовые дисплеи), графопостроители, печатающие устройства, экранные пульта и т. п.

Информационное обеспечение АСУ (ИО АСУ) — совокупность реализованных решений по объемам, размещению и формам организации информации, циркулирующей в АСУ в процессе ее функционирования. Основными составляющими подсистемы информационного обеспечения АСУ являются унифицированная система документации, машинные массивы информации, система классификации и кодирования.

Унифицированная система документации представляет собой рационально организованный комплекс взаимосвязанных документов, отвечающий единым правилам и требованиям, вытекающим из необходимости машинной обработки информации. Эта документация содержит информацию, необходимую для решения задач АСУ на базе ЭВМ и экономико-математических методов, и составляет основу внемашиной информационной базы АСУ.

Машинные массивы информации содержат необходимые исходные и справочные данные, а также нормативные данные для решения задач АСУ и составляют основу внутримашинной информационной базы АСУ.

Системы классификации и кодирования — важная составная часть методов и средств, связывающая документы и машинные массивы. На базе систем классификации и кодирования создается информационно-поисковый аппарат — один из важнейших элементов информационного обеспечения АСУ.

В целом информационное обеспечение объединяет данные в различных формах, средства формализованного описания данных, программные средства обработки данных, организационные принципы создания и ведения информационных массивов. Средства формализованного описания данных позволяют идентифицировать данные в информационных массивах и организовать эффективный доступ к ним. Помимо систем классификации и кодирования к средствам формализованного описания данных относятся информационные языки, предназначенные для описания запросов к массивам данных и ответов на эти запросы.

Программные средства обработки данных предназначены для контроля информации, вводимой в ЭВМ, ее хранения и накопления, внесения требуемых изменений в базу данных.

Организационные принципы создания и ведения внутримашинной информационной базы содержат методы и способы организации, хранения, изменения информационных массивов и доступа к ним.

Подсистема информационного обеспечения призвана накапливать, постоянно обновлять и хранить всю информацию, необходимую для решения задач в АСУ, и выдавать ее по запросам пользователей системы. Это определяет основные требования к подсистеме: достаточно полное отображение состояния объекта в любой момент времени; простой и быстрый доступ к информационной базе; избыточность данных; высокая эффективность методов и средств сбора, хранения, обновления, поиска и выдачи информации; возможность развития информационной базы за счет ее расширения и совершенствования методов обработки информации.

Огромные объемы информации, перерабатываемые в современных АСУ, требование коллективного характера внутримашинной информационной базы, пригодной для разнообразного применения в пределах решаемых задач, требование избыточности информационной базы и ее независимости от прикладных программ пользователей АСУ привели к созданию и развитию информационного обеспечения на основе автоматизированных банков данных (АБД). АБД можно определить как совокупность базы данных (БД) и системы управления базой данных (СУБД). БД представляет собой совокупность взаимосвязанных, хранящихся вместе данных, организованных так, что обеспечивается минимальная избыточность данных, возможность организации разнообразных структур, независимость организации данных от прикладных программ. СУБД — это программно-логический аппарат, организованный как пакет прикладных программ, предназначенный для эффективного управления базами данных. Подробно этот вопрос рассматривается в курсе «Базы и банки данных». Здесь же отметим, что организация внутримашинной информационной базы на основе банков данных — наиболее прогрессивная форма создания и развития информационного обеспечения АСУ.

Важной составной частью информационного обеспечения является нормативная база АСУ, содержащая нормативно-справочную

информацию в виде норм, нормативов, условно-постоянных справочных и учетных показателей. Нормативная база организуется в виде документов и машинных носителей.

Математическое обеспечение АСУ (МО АСУ) — совокупность математических моделей, методов и алгоритмов обработки информации в автоматизированной системе управления.

Важнейшей составляющей подсистемы математического обеспечения являются математические модели процессов и объектов управления. В сущности АСУ представляет собой взаимоувязанный комплекс экономико-математических моделей и процессов вместе с методами и средствами их анализа. Характер моделей определяется характером решаемых задач управления. Формально в большинстве случаев модель конструируется в виде целевой функции и системы ограничений. Решение задачи в таких случаях сводится к поиску экстремума целевой функции при соблюдении ограничений. Способы решения таких задач базируются на методах математического программирования, значительная часть которых рассмотрена в гл. 7.

В частных случаях модели процессов и объектов могут быть сведены к аналитическим детерминированным моделям. Примером такой целевой функции может служить модель вида

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min \quad (9.1)$$

при ограничениях

$$\sum_{i=1}^m x_{ij} = b_j; \quad \sum_{j=1}^n x_{ij} = x_i \leq a_i, \quad i = \overline{1, m}; \quad j = \overline{1, n}; \quad (9.2)$$

$$x_i \geq 0; \quad x_{ij} \geq 0.$$

Задача, формализованная в виде (9.1), решается методами линейного программирования. К такой целевой функции может быть сведена модель транспортно-производственной задачи. В этом случае обозначения в (9.1) имеют следующий смысл: m — число объектов производства продукта; n — число пунктов потребления продукта; a_i — мощность i -го объекта производства; b_j — объем продукции, потребляемой j -м заказчиком; c_{ij} — стоимость производства и доставки единицы продукта из i -го пункта производства в j -й пункт потребления; x_{ij} — объем продукта, поставляемого i -м производством j -у потребителю.

Ограничения (9.2) в рассматриваемом примере обусловлены требованием максимальной эффективности производства и потребления продукта.

Аналитические детерминированные модели в составе математического обеспечения АСУ в большинстве случаев не могут отражать сложности решаемых задач. Более реальными являются комбинированные модели и комбинированные методы моделирования, основанные на одновременном использовании двух критериев выбора решений — формализованного и эвристического.

Экономико-математические модели производственных объектов могут строиться на основе производственных функций, которые

представляют собой функциональные зависимости между различными факторами системы управления, определяемые статистическими или корреляционными методами. Производственные функции устанавливают соотношение между выпуском продукции и используемыми ресурсами:

$$F(Y, X, C) = 0, \quad (9.3)$$

где $Y = \{y_1, y_2, \dots, y_i, \dots, y_n\}$ — совокупность показателей выпуска продукции; $X = \{x_1, x_2, \dots, x_j, \dots, x_m\}$ — совокупность показателей ресурсов (рабочая сила, основные и оборотные фонды, сырье и т. п.); $C = \{c_1, c_2, \dots, c_r, \dots, c_k\}$ — k -мерный вектор.

Сложность задач управления в АСУ заключается в том, что далеко не все факторы, влияющие на объекты и процессы управления, могут быть описаны количественно. Поэтому наиболее типичными задачами являются такие, для которых характерна постановка и решение в условиях неполной информации. В этом смысле важное значение имеют методы решения, основанные на принципах стохастического программирования.

Программное обеспечение (ПО) — совокупность программ для реализации задач АСУ с помощью комплекса технических средств. ПО АСУ состоит из двух частей: 1) общего программного обеспечения — представляющего собой совокупность программ, ориентированных на широкий круг пользователей для организации вычислительного процесса и решения часто встречающихся задач обработки информации; 2) специального программного обеспечения — включающего комплекс программ для реализации функций конкретной системы управления. Первая часть ПО создается разработчиками вычислительной системы или ЭВМ, вторая — разработчиками АСУ.

Одним из важнейших принципов разработки ПО АСУ является принцип программной совместимости, призванный обеспечить обмен программами в условиях работы АСУ различных уровней.

Общее программное обеспечение является универсальным и практически не зависит от области применения ЭВМ. Составляющими общего ПО являются комплексы программ: технического обслуживания вычислительной системы; операционной системы; системы автоматизации программирования; системный диспетчер; библиотека стандартных программ и подпрограмм общего назначения.

Программы технического обслуживания ЭВМ обеспечивают правильное функционирование вычислительной системы, используя, в частности, тестирование. Операционная система организует вычислительный процесс, распределяя все ресурсы ЭВМ: машинное время процессора; емкость оперативной и внешней памяти; работу устройств ввода-вывода. Средства автоматизации программирования включают трансляторы с алгоритмических языков программирования, стандартные программы и подпрограммы основных математических процедур, программы решения типовых задач. Системный диспетчер содержит программы, обеспечивающие функционирование вычислительной системы в ритме работы объекта

управления АСУ, связь ЭВМ с периферийными устройствами и абонентами, порядок выполнения рабочих программ. Библиотека стандартных программ объединяет стандартные программы информационного обеспечения, обеспечивающие организацию информационных массивов, ввод информации, формирование машинной формы документов и т. п.

Специальное ПО формируется в виде совокупности пакетов прикладных программ (ППП). Каждый ППП представляет собой комплекс взаимосвязанных программ, предназначенный для реализации конкретной функции АСУ и настраиваемый при конкретном применении.

Лингвистическое обеспечение АСУ — совокупность языковых средств для формализации естественного языка, построения и сочетания информационных единиц при общении персонала АСУ с ЭВМ. Необходимость использования нескольких языков программирования объясняется разнообразием задач, решаемых в АСУ, и степенью приспособления того или иного языка к конкретной задаче. В этом смысле языки программирования можно разбить на три группы: 1) машинные (МЯ); 2) машинно-ориентированные (МОЯ); 3) проблемно-ориентированные (ПОЯ).

МЯ — это машинные коды. В состав МЯ входит перечень операций, свойственный конкретной ЭВМ, с их числовыми кодами. МЯ применяются в редких случаях — когда требуются программы высшего качества при минимальном времени их реализации. В остальных случаях МЯ не применяются из-за крайне низкой производительности процесса программирования и необходимости знания детальных особенностей ЭВМ.

МОЯ — группа языков, также ориентированных на конкретную ЭВМ или тип ЭВМ. Однако МОЯ содержит элементы автоматизации процесса распределения памяти ЭВМ. Например, в мнемокоде, одном из МОЯ, цифровые коды операций машинного языка ЭВМ заменены буквенными или буквенно-цифровыми кодами. В другом МОЯ — автокоде — часто используемые совокупности машинных команд объединены в макрокоманды, что повышает производительность процесса программирования. Примером широко применяемого машинно-ориентированного языка является язык АССЕМБЛЕР, который использует набор макрокоманд, позволяющий реализовать функции операционной системы ЕС ЭВМ.

ПОЯ — группа языков программирования, ориентированных на особенности решаемых задач, а не на ЭВМ. По структуре ПОЯ ближе к естественному языку, чем к машинному, поэтому программирование на этих языках не требует детального знания особенностей конкретной вычислительной машины и отличается высокой производительностью. Для перевода ПОЯ на язык конкретной машины в составе программного обеспечения ЭВМ имеются *трансляторы* — программы перевода, поэтому в распоряжении пользователей ЭВМ в АСУ есть несколько языков программирования высокого уровня, к которым относятся ПОЯ: ФОРТРАН, КОБОЛ, АЛГОЛ, РПГ, ПЛ/1 и др.

ФОРТРАН (*FORT*RAN — *FOR*Mula *TRAN*slator) используется

для программирования задач научно-технического характера. Язык базируется на алфавите естественного языка, содержит развитые средства связи между автономными частями программы — подпрограммами, имеет удобные средства отладки программ, эффективные средства использования устройств ввода-вывода. Трансляторы с ФОРТРАНа имеют средства выявления синтаксических ошибок в исходной программе.

КОБОЛ (*COBOL — Common Business Oriented Language*) используется для программирования задач организационно-экономического характера, так как содержит эффективные средства описания больших информационных массивов. Имеет развитую библиотеку, программы которой реализуют отдельные операторы языка, преобразование данных из одного формата в другой, арифметические операции и специальные функции.

АЛГОЛ (*ALGOL — ALGOritmic Language*) ориентирован на задачи численного анализа. Блочная структура языка, динамическое распределение памяти, развитый аппарат вызова процедур, значительное число библиотек программ сделали АЛГОЛ базовым языком для многих других языков программирования.

РПГ (*RPG — Report Programm Generator*) ориентирован на решение задач, не требующих большого объема вычислений, но имеющих большой объем операций ввода-вывода данных. На языке РПГ удобно программировать алгоритмы решения экономических задач, содержащих значительное количество операций сложения, вычитания, умножения, деления, вычисления итогов, обработки таблиц и т. п.

ПЛ/1 (*PL/1 — Programming Language/One*) — универсальный язык программирования, пригодный для программирования широкого круга задач: научно-технических, организационно-экономических, информационно-логических. В ПЛ/1 входят элементы языков ФОРТРАНа, АЛГОЛа, КОБОЛа, вместе с тем он содержит ряд новых элементов, что в целом сделало его новым универсальным языком программирования высокого уровня. Язык содержит набор средств, позволяющих эффективно описывать вычислительные процессы, процедуры обработки данных и символьной информации, процессы моделирования, отражать процессы решения задач в реальном масштабе времени. Программирование для ЭВМ на языке ПЛ/1 позволяет повысить эффективность использования всей вычислительной системы в целом.

Выше дана характеристика языков программирования прикладных задач АСУ. Кроме этого, в составе лингвистического обеспечения вычислительной системы содержатся информационные языки, используемые для обработки данных: запроса к базе данных; формирования различных логических отношений между элементами структур базы данных; передачи информации в базу данных; управления внешними устройствами. К числу таких языков относятся: языки описания данных (ЯОД) для описания структуры и содержания базы данных; язык команд (ЯК), используемый пользователем для взаимодействия с базой данных; язык управления внешними устройствами (ЯУВ). Перечисленные информационные

языки являются средствами СУБД — систем управления базами данных.

В заключение заметим, что математическое, программное и лингвистическое обеспечения АСУ тесно увязаны между собой и часто рассматриваются как единая система программно-математического обеспечения. Роль этих подсистем весьма велика. Если техническое обеспечение аналогично станочному парку некоторого производства, информационное — сырьевым ресурсам, то программно-математическое обеспечение можно рассматривать как совокупность инструментальных средств, от качества которых в значительной степени зависит качество конечной продукции, что применительно к АСУ означает достоверность результатов решения задач управления и производительность вычислительных операций. Развитие АСУ характеризуется непрерывным возрастанием удельной стоимости программно-математического обеспечения, характеризующего в значительной мере эффективность функционирования АСУ в целом. В настоящее время средства программно-математического обеспечения оцениваются в 70 % и выше общей стоимости АСУ.

Организационное обеспечение АСУ — совокупность документов, регламентирующих деятельность персонала АСУ в условиях ее функционирования.

Правовое обеспечение АСУ — совокупность правовых норм, регламентирующих правоотношения при функционировании АСУ и юридический статус результатов ее функционирования.

9.3. ОСНОВЫ РАЗРАБОТКИ АСУ

Создание АСУ представляет собой сложный и длительный процесс, включающий в себя научно-исследовательские, проектные, строительные, монтажные, наладочные и другие работы. Процесс создания АСУ разбивается на стадии и этапы. *Стадия* — одна из последовательно выполняемых частей процесса создания АСУ, установленных нормативными документами. Каждая стадия заканчивается получением заданного результата, которым является назначенный комплект документации или определенный объем внедрения автоматизированного управления. *Этап* составляет часть стадии, установленную нормативными документами на основе рационального планирования работ по созданию АСУ.

Введены три стадии разработки АСУ: 1) предпроектная; 2) разработка проектов; 3) ввод в эксплуатацию.

Предпроектная стадия включает в себя этапы разработки технико-экономического обоснования (ТЭО) и технического задания (ТЗ).

Важнейшей частью работ на предпроектной стадии создания АСУ является исследование существующей традиционной системы управления. Цель этого исследования заключается в том, чтобы вынести суждение о целесообразности перевода объекта на автоматизированное управление и наметить основные требования к будущей АСУ, если будет принято решение о ее создании. Важно под-

черкнуть, что уже на стадии обследования традиционной системы управления могут быть обнаружены значительные резервы повышения эффективности управления.

При исследовании объекта управления выясняется состояние его производственной деятельности, его организационная структура и система управления, взаимодействие с поставщиками и потребителями, связи с научно-исследовательскими и проектными организациями. Весьма важной и трудоемкой составной частью исследования является изучение потоков информации на основе изучения и анализа форм документов, техники их заполнения и обработки. Трудоемкость этой работы определяется множеством вопросов, которые нужно решить при анализе документооборота системы управления: назначение каждого документа; количество экземпляров; обязательные реквизиты и показатели документов; значимость каждого показателя и правила их формирования; периодичность составления документов; маршрут движения документа в пределах исследуемой системы и др.

Для упрощения и сокращения процесса исследования объекта управления широко используются методы и средства формализованного представления отдельных этапов исследования. В частности, используются специальные формы — спецификации определения систем управления: бланки описания организаций; структурная схема организации; таблица функций организации и ее подразделений; форма характеристик задач; бланк описания информационных потоков. Результирующим документом на этапе изучения структуры традиционной системы управления, целей ее функционирования и ограничений, а также информационных потоков в системе является обобщенная структурно-информационная временная схема (ОСИВС), на которой с помощью условных обозначений отражаются этапы и маршруты движения документов.

Для анализа информационных потоков и алгоритмов обработки информации используются специальные формы: анализ документов; анализ массивов; анализ процедур.

Дальнейшим шагом совершенствования этапа исследования традиционных систем управления является определение информационно-функциональных характеристик объектов управления с использованием ЭВМ, что призвано значительно сократить время обследования объектов и повысить качество результатов обследования. Автоматизированные системы анализа информационных характеристик объектов управления позволяют получить такие сведения, как полнота обследования, степень обоснованности показателей документов, схемы формирования показателей и др.

Разработка проектов АСУ содержит этапы разработки технического и рабочего проектов. В ряде случаев, когда накоплен определенный опыт проектирования тех или иных типов АСУ, возможно одностадийное проектирование, при котором разрабатывается единый технорабочий проект. С другой стороны, при значительных масштабах объекта и отсутствии опыта проектирования возможно трехстадийное проектирование, при котором, помимо ра-

бочего и технического проектов предварительно разрабатывается эскизный проект.

Технорабочий проект содержит следующие основные материалы: пояснительную записку; решения по техническому обеспечению; программную документацию; инструкции по обработке информации; должностные инструкции; описания постановки и решения задач управления; описание организационной структуры АСУ; организацию информационной базы; системы классификации и кодирования технико-экономической информации; расчет экономической эффективности с указанием источников эффективности и срока окупаемости. Состав и содержание документов технорабочего (и других) проектов определяются ГОСТом.

В стадию разработки проектов АСУ входит также подготовка объекта к вводу в эксплуатацию, включающая разработку документов на заявку технических средств, их монтаж и наладку, контрольную проверку программ и инструкций.

Ввод в эксплуатацию включает в себя завершение строительно-монтажных и пусконаладочных работ, опытную эксплуатацию АСУ и сдаточные испытания. Осуществляется постепенный переход от традиционной системы управления к автоматизированной.

Большие масштабы работ по проектированию и внедрению АСУ в народное хозяйство потребовали разработки промышленных методов создания АСУ, основанных на типовых элементах и типовых проектных решениях.

Типовые проектные решения (ТПР) АСУ — типовая проектная документация по программно-математическому и техническому обеспечению, позволяющая проектировать АСУ путем использования этой документации совместно с оригинальной проектной документацией, учитывающей специфику объекта управления. Уровень типизации особенно высок при разработке подсистем технического обеспечения АСУ, где он достигает 70 % и более. При разработке программно-математического обеспечения уровень типовых проектных решений составляет 10—50 %.

В целом по АСУ ТПР составляют от 30 до 70 %, что позволяет сократить сроки разработки АСУ и затраты на 15—30 %.

Разработка АСУ на базе ТПР сводится к выбору подходящего типового проекта решения и разработке проекта привязки ТПР к конкретному объекту управления. В проекте привязки устанавливаются отличия проектируемого объекта по экономическим показателям, информационным и функциональным характеристикам, организационной структуре, методам реализации функций управления, особенностям технологических процессов и т. п. Для учета этих отличий создаются проектные документы, составляющие оригинальную (не типовую) проектную документацию.

Типовые проектные решения строятся по модульному принципу на основе блочной структуры. При этом каждый из блоков реализует определенную часть задачи с использованием модуля-алгоритма и соответствующего ему программного модуля. Модуль-алгоритм и программный модуль выполняются в виде структуры,

характеризующейся значительной степенью законченности и пригодной для многократного использования при разработках АСУ на основе ТПР, а также во взаимодействии с оригинальной частью проекта системы управления.

Модульная структура ТПР, включение в состав ТПР нескольких вариантов модулей для реализации задач, отличающихся некоторыми показателями, возможность расширения фонда модулей обуславливают широкую область применения ТПР.

Структура ТПР содержит три класса типовых проектных решений: «Задача»; «Техника»; «Персонал».

«Задача» — класс ТПР, включающий в себя комплекс алгоритмов и программ, реализующих определенные функции управления в различных функциональных подсистемах АСУ. К ним относятся подсистемы: управления технической подготовкой производства; технико-экономического планирования; оперативного управления; управления материально-техническим снабжением; управления сбытом и реализацией продукции; бухгалтерского учета.

«Техника» — класс ТПР, включающий типовую проектную документацию по составу, размещению и использованию технических средств АСУ.

«Персонал» — класс ТПР, содержащий типовую проектную документацию, определяющую действие персонала АСУ в обычных и аварийных режимах работы.

Дальнейшим совершенствованием процесса разработки АСУ является использование систем автоматизированного проектирования (САПР АСУ). Автоматизированное проектирование АСУ опирается на методы автоматизированного анализа информационно-функциональных характеристик объектов управления, автоматизированного проектирования без данных, вычислительного процесса, программирования, средств формирования проектной документации. Основной проблемой, требующей решения и определяющей эффективность САПР АСУ, является необходимость разработки системы экономико-математических моделей функционирования различных объектов управления.

Примером САПР АСУ может служить разработанная в Ленинградском объединении «Ленэлектронмаш» «Информационная система управления производством», представляющая собой пакет прикладных программ для реализации обобщенных типовых решений основных задач управления производством. Наиболее полно в системе решаются задачи оперативного управления производством, а также ряд функций управления технико-экономическим планированием и материально-техническим снабжением.

Из зарубежных разработок широкую известность получила система *PICS (The Production Information and Control System)*, представляющая собой комплект пакетов прикладных программ, реализующий формирование информации, необходимой для управления подготовкой производства. Гибкая база данных обеспечивает широкую область применения системы.

Основное назначение САПР АСУ — автоматическая разработка проблемно-ориентированного математического обеспечения и

рабочего программного обеспечения на базе анализа информационно-функциональных характеристик объектов управления с автоматической выдачей проектной документации на разрабатываемую АСУ.

Использование ТПР и САПР АСУ призвано существенно повысить эффективность и темпы разработки автоматизированных систем управления в различных отраслях народного хозяйства.

9.4. КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

АСУ, как любая сложная система, функционирует в условиях, характеризующихся высокой степенью неопределенности. Объект управления и вся система управления подвержены воздействиям, характер и параметры которых в ряде случаев трудно предсказать. В связи с этим важное значение имеют методы и приемы выбора оптимальных или рациональных решений в неопределенной ситуации, когда невозможно достаточно определенно судить о характере возможных воздействий на систему.

Приведем несколько общих критериев рационального выбора вариантов решений из множества возможных альтернатив. Критерии основаны на анализе матрицы возможных воздействий и альтернатив решений в исследуемой системе (табл. 9.1).

Матрица содержит: A_i — альтернативы, т. е. варианты состояний, один из которых необходимо выбрать; S_j — возможные воздействия, т. е. состояния входных сигналов; a_{ij} — оценки выбранной альтернативы, характеризующие доходы или убытки, возникающие от того, что при выбранной альтернативе возможны отклонения от соответствующего состояния входных сигналов.

Альтернативы в описанных условиях могут выбираться по одному из критериев, описанных ниже.

Таблица 9.1

A	S					
	S_1	S_2	...	S_j		S_m
A_1	α_{11}	α_{12}	.	α_{1j}		α_{1m}
A_i	α_{i1}			α_{ij}		α_{im}
A_n	α_{n1}			α_{nj}		α_{nm}

Критерий Лапласа определяет выбор альтернатив по условию

$$\max_i \left| \frac{\sum_j a_{ij}}{m} \right|.$$

Этот критерий исходит из равной вероятности всех состояний воздействий S_j на систему. Стратегия поведения лица, принимающего решения (ЛПР), в этом случае аналогична поведению военачальника, который, не сумев определить направление ожидаемого главного удара противника, распределил свои силы равномерно по всему фронту.

Критерий Вальда. При выборе альтернативы по этому критерию в каждой строке матрицы определяется минимальный результат и выбирается строка с наибольшим из этих значений:

$$\max_i \left| \min_j a_{ij} \right|.$$

По этому критерию ЛПР выбирает стратегию, гарантирующую максимальное значение наихудшего выигрыша (стратегия критерия максимина).

Критерий Гурвица выражается в формуле

$$\max_i [\alpha (\max_j a_{ij}) + (1 - \alpha) (\min_j a_{ij})],$$

где α — коэффициент оптимизации, назначаемый субъективно из условия $0 < \alpha < 1$. Эффективность выбора альтернативы зависит от того, насколько удачно выбрано значение α . При $\alpha = 1$ имеет место расчет на самую благоприятную ситуацию, при $\alpha = 0$ критерий Гурвица превращается в критерий Вальда, гарантируя наилучший результат в наихудших условиях воздействий на систему.

Критерий Сэвиджа исходит из «минимума сожаления» между выбранной и наиболее благоприятной альтернативой, соответствующей истинной ситуации:

$$\max_i [\min_j b_{ij}] = \max_i \{ \min_j [a_{ij} - (\max_i a_{ij})] \}.$$

Степень «сожаления» характеризуется величиной $b_{ij} = a_{ij} - (\max_i a_{ij})$, которая принимает отрицательные или нулевые значения.

Критерий благоприятного в среднем решения основан на определении математического ожидания дохода, соответствующего выбранной стратегии поведения, по известным вероятностям $P(S_j)$ состояний воздействий на систему. Вероятностные характеристики $P(S_j)$ могут быть получены, например, в результате изучения статистических сведений и экстраполяции этого результата на ближайшее будущее.

Используя рассматриваемый критерий, ЛПР вырабатывает альтернативу, обеспечивающую наибольший ожидаемый доход:

$$\max_i [\sum_j P(S_j) a_{ij}].$$

С необходимостью принимать решения в условиях неопределенности приходится сталкиваться на всех стадиях и этапах раз-

работки и эксплуатации АСУ. При этом формализация процессов принятия решения в форме математических моделей в большинстве случаев либо невозможна, либо может дать лишь весьма общий результат, не учитывающий важные особенности конкретной обстановки, в которой требуется вырабатывать решения. В связи с этим при разработке и эксплуатации АСУ важную роль играют неформальные методы выдвижения творческих идей в процессе решения возникающих проблем. Одним из популярных методов такого рода является метод «мозгового штурма», стимулирующий творческое мышление. Для реализации метода комплектуется группа из 6—10 человек, в которую входят представители научно-исследовательских, конструкторских и производственных подразделений во главе с председателем — лицом, имеющим опыт применения метода. Наряду со специалистами, глубоко занимающимися рассмотренной проблемой, в группу включаются 1—2 человека, не знакомых с проблемой, являющихся специалистами в других областях науки и техники.

Метод реализуется в два этапа. На первом этапе предлагается высказывать любые идеи, включая такие, которые опытным специалистам покажутся бессмысленными. При этом для поощрения к выдвижению большого количества идей критика высказываемых идей запрещается, допускается их уточнение и комбинирование. На втором этапе все выдвинутые идеи внимательно анализируются опытными специалистами-экспертами и оцениваются с использованием заранее разработанных таблиц критериев. Идеи, в наибольшей степени отвечающие всем критериям, принимаются к реализации.

9.5. ГИБКИЕ АВТОМАТИЗИРОВАННЫЕ ПРОИЗВОДСТВА

Гибкое автоматизированное производство (ГАП) — это производственный комплекс, ориентированный на выпуск законченной товарной продукции и основанный на совокупности автоматических линий, участков, цехов, заводов, объединенных в интегрированные производственные системы.

Основное достоинство ГАП — гибкость, под которой понимается способность оперативно перестраиваться на выпуск новой продукции. Это достигается: использованием соответствующих методов управления, обеспечивающих оперативную перестройку производства; использованием комплексов программного управления, реализуемых на ЭВМ различных классов: микроЭВМ, мини-ЭВМ, больших ЭВМ; созданием и внедрением новых методов, средств, устройств, технологий и т. п., обеспечивающих высокую надежность всего комплекса в условиях безлюдной технологии отдельных его участков.

ГАП представляет собой высший этап комплексной автоматизации производства на основе ЭВМ и экономико-математических методов. Методология ГАП содержит в себе огромные резервы повышения производительности труда и уровня организации производства. В перспективе целые производства, построенные на ос-

новые принципы ГАП, могут работать по безлюдной технологии. Однако, несмотря на высокую степень автоматизации, ГАП не является автоматической системой, а относится к типу человеко-машинных систем. Эта предпосылка весьма важна, так как определяет принципиальный подход к разработке и проектированию ГАП, методы и средства разработки. Являясь автоматизированной, а не автоматической системой, ГАП должно разрабатываться на базе принципов системного подхода на основе базовых моделей процесса управления. Анализ таких математических моделей позволяет определить три основных режима функционирования ГАП: контроль, поиск решения, реализация решения.

В *режиме контроля* качество функционирования ГАП контролируется с использованием комплекса математических моделей, на базе которых результат контроля может быть представлен в наглядной форме. При отклонении хода управляемого процесса от запланированного с помощью моделей определяется время t , в течение которого требуется выявить и устранить причину отклонения заданного режима. Качество контроля определяется качеством моделей и полнотой информации о состоянии системы.

В *режиме поиска решений* ГАП переходит, если на этапе контроля устанавливается необходимость перестройки режима работы системы. Поиск оптимального или приемлемого решения осуществляется на базе глобальной математической модели всего производственного процесса. Окончательное решение принимается ЛПР (или лицами).

В *режиме реализации решения* система перестраивается по плану, выработанному на этапе поиска решения, причем каждая операция по реализации принятых решений осуществляется под автоматизированным контролем.

Описанные режимы функционирования ГАП полностью характеризуют эту систему как человеко-машинную, как автоматизированную систему управления.

В целом ГАП укрупненно можно представить как иерархическую трехуровневую систему, в которой на базе электронно-вычислительных машин и экономико-математических методов осуществляется комплексная автоматизация всего производственного процесса.

При этом на нижнем уровне автоматизируются простейшие рабочие операции, например станочная обработка деталей, сварка, резка и т. п. По существу, на первом уровне обеспечивается автоматическое управление технологическим процессом с оптимизацией режимов. Элементной базой автоматизации на этом уровне являются роботизированные комплексы, управляемые с помощью микроЭВМ и микропроцессорных устройств.

На втором уровне реализуется автоматизированное организационно-технологическое управление. На этом уровне синхронизируется работа модулей обработки, контроля качества, транспортно-накопительных и других систем на базе автоматизированного диспетчера с использованием динамической модели модуля складирования и комплектации. Управление второго уровня осуществ-

вляется на базе терминальных станций обработки технико-экономической информации, связанных с центральной ЭВМ.

На третьем уровне реализуется оперативно-производственное управление на базе недельного и сменно-суточного планирования, учета и контроля. Управление на третьем уровне осуществляется на базе мини-ЭВМ, связанных в единый управляюще-вычислительный комплекс с центральной ЭВМ. Подсистемы ГАП, обеспечивающие третий уровень управления, представляют собой автоматизированные системы управления производством и автоматизированные системы технологической подготовки производства.

Перестройка ГАП на выпуск новых партий изделий осуществляется программно-аппаратными средствами управляющей части с минимальным изменением основного технологического оборудования. Степень гибкости ГАП определяется степенью приспособляемости производственной системы к динамическим и стохастическим изменениям, вытекающим из общей производственной программы. ГАП должно быстро переналаживаться на выпуск изделий различной номенклатуры, определенной для каждого производства. В гибкой производственной системе (ГПС) переход к новой производственной программе осуществляется с сохранением числа и вида элементов системы и их связей.

Отвечающая указанным требованиям функциональная схема ГАП представлена на рис. 9.1.

Выпуск заданной продукции осуществляется на робототехнических участках механической обработки (РУМО) и сборки (РУС) в соответствующих роботизированных ячейках (РЯ), входящих в транспортно-накопительную систему (ТНС). Исходные материалы на РУМО поступают с автоматизированного склада заготовок (АСЗ) под управлением управляюще-вычислительного комплекса (УВК). Необходимый инструмент и технологическая оснастка на РУМО так же, как и на РУС, поступают по транспортной системе (ТС), связывающей эти участки с автоматизированным складом хранения инструмента (АСХИ) и автоматизированным складом хранения технологической оснастки (АСХТО).

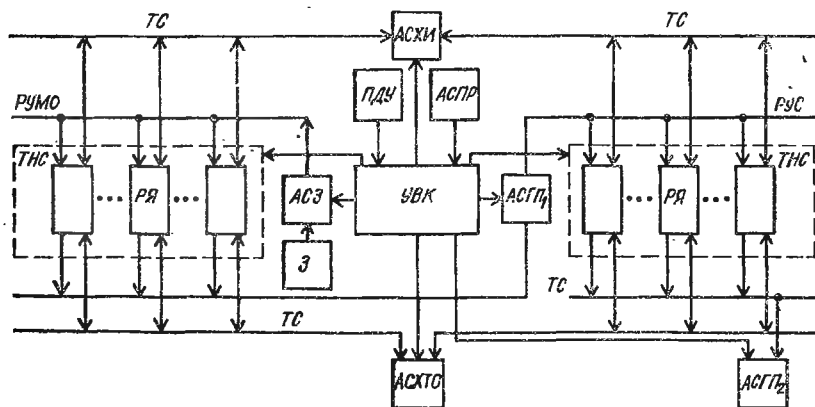


Рис. 9.1. Функциональная схема гибкого автоматизированного производства

С участка механической обработки изделия поступают на автоматизированный склад готовой продукции (АСГП₁), откуда по командам УВК подаются на участок сборки (РУС). Готовая продукция с РУС поступает на АСГП₂ под управлением УВК.

Автоматизированные системы проектирования (АСПР) включают в себя системы: автоматизированного проектирования выпускаемой продукции — изделия (САПРИ); автоматизированной разработки технологических процессов изготовления продукции (САПТ); автоматизированную систему организационно-экономического планирования (АСОЭП); автоматизированного проектирования технологической подсистемы (САПТПС), в которую входят проектирование робототехнических участков механической обработки и сборки.

Все операции по доставке исходных заготовок (З), механической обработке деталей, сборке изделий, доставке требуемого инструмента и технологической оснастки, направлению продукции на склады и т. д. осуществляются под управлением УВК на базе специального математического обеспечения ГАП. Сигналы управления на УВК поступают из поста диспетчерского управления (ПДУ) или АСПР. Основу технологической подсистемы ГАП составляет оборудование с числовым программным управлением (ЧПУ). Модуль с ЧПУ подразделяется на две автономные части — управляющую и исполнительную. В управляющую часть загружается программа, определяющая управляющие воздействия, обеспечивающие автоматический режим работы технологической подсистемы. Исполнительная часть обеспечивает рабочие операции процесса обработки и сборки.

Как было отмечено выше, управление гибким производством осуществляется на трех уровнях. Каждому уровню соответствует свой комплекс задач и состав технических средств (рис. 9.2).

Автоматизированные системы технологической подготовки производства (АСТПП) и управления производством (АСУП) реализуют третий уровень управления, техническую основу которого составляют автоматизированные рабочие места (АРМ) и мини-ЭВМ. Обе эти подсистемы связаны с центральной ЭВМ, составляющей основу системы управления гибким производством (СУГП). АСТПП, АСУП и СУГП вместе образуют управляющий вычислительный комплекс (УВК) гибкой производственной системы (ГПС), реализующий программы управления планированием и технологической подготовкой производства. СУГП обеспечивает АСУП необходимыми данными для оперативного планирования и учета производства. АСТПП обеспечивает программы работы гибкого производства для всей номенклатуры изделий. Реализация этих программ позволяет разработать маршруты движения обрабатываемых объектов между отдельными участками производства, модулями обработки и сборки, определить состав инструмента на каждом технологическом участке, автоматизировать составление технологического процесса на изделие.

Центральная ЭВМ составляет основу второго уровня управления, а связанные с ней мини- и микроЭВМ через устройство связи

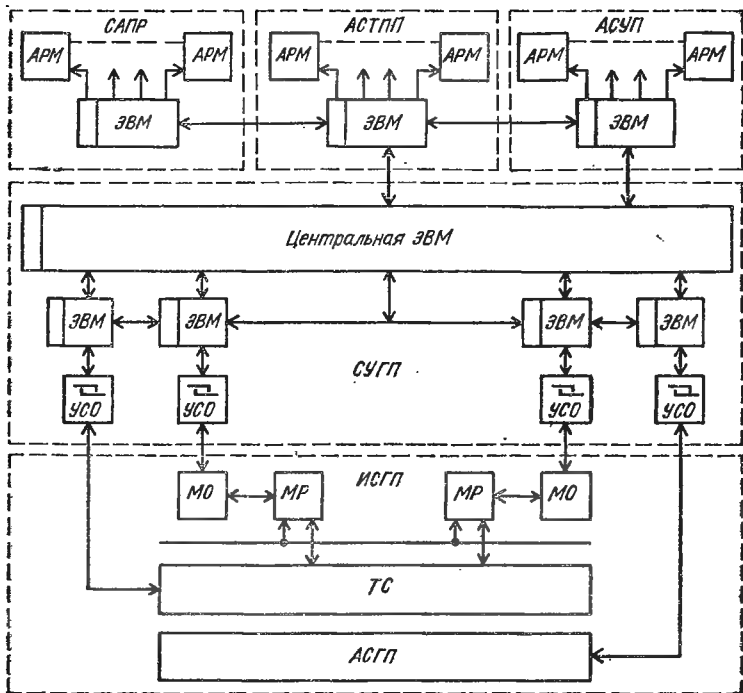


Рис. 9.2. Схема управления гибким автоматизированным производством

с объектом (УСО) осуществляют управление первым уровнем — модулями обработки (МО) и модулями робототехники (МР). Эти модули, транспортная система (ТС) и автоматизированный склад готовой продукции (АСГП) вместе образуют исполнительную систему гибкого производства (ИСПП). Система автоматизированного проектирования (САПР), связанная с АСТПП, обеспечивает автоматизированное проектирование изделий, предназначенных для выпуска гибким производством, а также для оперативного проектирования технологии, оснастки и инструмента.

Результаты проектирования получаются в виде информации на машинных носителях, используемых для управления технологическим оборудованием при изготовлении изделий. Таким образом, отпадает необходимость такой громоздкой и дорогостоящей работы, как выполнение чертежей и другой бумажной документации.

Управление гибким производством в целом обеспечивается на базе совокупности программ, реализуемых комплексом ЭВМ, сопряженных между собой в единый управляюще-вычислительный комплекс с помощью линий передачи информации. Набор программ обеспечивает обмен данными между ЭВМ в каждом уровне управления и между всеми уровнями.

Программное управление каждой подсистемы и всего гибкого производства базируется на модульной основе. Библиотека модульных программ обеспечивает формирование пакетов прикладных

программ (ППП) для решения конкретных задач в ГПС. В результате управление ГПС в значительной степени сводится к загрузке программ в ЭВМ, обеспечивающих реализацию модулей исполнительской подсистемы в соответствии с планом выпуска изделий, заданной технологией и темпом работы оборудования.

Реализация программных модулей тесно увязывается с временной диаграммой функционирования ГПС и текущим состоянием ее элементов.

Весь комплекс программного обеспечения ГАП делится на две части — обслуживающие программы и функциональные (технологические).

Обслуживающие программы составляют операционную систему ЭВМ и предназначены для управления ресурсами ЭВМ. Примерами частных функций, обеспечиваемых обслуживающими программами, являются управление загрузкой программ в память ЭВМ, контроль работы КТС вычислительной системы, управление обменом данными между модулями и др.

Функциональные программы определяют порядок работы управляемых объектов, вытекающий из заданной технологии. Загруженные в память ЭВМ функциональные программы настраивают элементы ГПС на выполнение заданных функций.

Таким образом, ГАП характеризуется как многоуровневая система программного управления, реализуемая комплексом ЭВМ с использованием наборов программ, организованных в программные модули. Именно это в значительной степени обеспечивает главное достоинство ГАП — оперативное переналадивание производства на выпуск изделий различной номенклатуры с минимальными изменениями технологического оборудования или вообще без изменения его состава.

А

АБД — автоматизированный банк данных 275
 Автоколебания 85, 93
 Автоматизация проектирования 261
 Автоматизированное управление 14, 15
 Автоматизированные системы управления 14, 279
 Автоматическое регулирование 12
 — управление 12
 Алгоритм цикла проектирования 264
 Алфавитно-цифровые печатающие устройства 21
 Амплитудно-фазовые характеристики 55
 Анализ импульсной системы 106
 — модели на чувствительность 206
 — систем 66, 82
 — статических режимов 32
 Аналого-цифровой преобразователь 170, 171
 Аналоговые электронно-вычислительные машины (АВМ) 151
 Астатическая система 28
 Астатическое регулирование 28

Б

Базис опорного плана (решения) 203
 Базы данных 275
 Банки знаний 156
 Банк промежуточных результатов 279
 Бифуркационные параметры 91
 Блок данных 275
 Блочная структура языка 287
 Булев параметр 36
 Булевы переменные 36
 Буферная память 276
 Быстродействие 24
 Быстродействующие устройства памяти 157

В

Вариационное исчисление 123
 Вектор возмущения 10, 122
 — входных воздействий 10
 — выходных величин 10, 122
 — начального состояния объекта 122
 — состояния объекта 122
 — управления 10, 122
 Векторное представление модели 187
 Вещественная частотная характеристика 78, 80
 — — замкнутой системы 80
 Воздействия возмущающие 10
 — входные 10
 — типовые 33
 — управляющие 10
 Внешняя память ЭВМ 157
 Время переходного процесса 33

Г

Гибкие автоматизированные производства 297
 Гибридная вычислительная система 155
 Гипербола Вышнеградского 74
 Глобальный минимум целевой функции 240
 Годограф вектора 69
 — Михайлова 69
 Градиент 135
 Граница устойчивости 111
 Граф 210
 Графические методы 99
 Графопостроители 276

Д

Двоичная система счисления 157
 Двойственная задача 197, 198
 Двойственность в линейном программировании 197
 Двойственный симплекс-метод 202

Декада 58
 Дельта-функция 81
 Диаграмма Вышнеградского 73
 Диалоговый режим 262
 Динамические режимы 29, 33
 — свойства САУ 33
 D-разбиение 75
 Диск магнитный 157
 Дискретные системы 99
 Дисплеи 276
 — графические 276
 — текстовые 276
 Дифференциальные уравнения системы 52
 Дуальный процесс 155

Е

Естественные ограничения 119
 Единичная трапеция 79

Ж

Желаемая ЛАЧХ 81

З

Задача в условиях риска 251
 — календарного планирования 182
 — комбинаторной оптимизации 228
 — линейного программирования 182
 — о назначениях 182
 — оперативного стохастического программирования 253
 — определения кратчайшего пути 223
 — перспективного стохастического программирования 253
 Задающий элемент 12
 Закон управления 25
 — — интегральный 25
 — — пропорциональный 25
 Запас устойчивости по фазе 72
 Звено 29
 — аperiodическое 41
 — дифференцирующее 46
 — — второго порядка 47
 — — идеальное 46
 — — первого порядка 47
 — интегрирующее 45
 — колебательное 43
 — неустойчивое аperiodическое 42
 — — колебательное 43
 — усилительное 45
 Зона нечувствительности нелинейного элемента 93

И

Идентификатор процесса 147
 Изображающая точка 87
 Изображение функции 35
 Импульсная переходная функция 81

Импульсные системы 100
 Импульсный элемент 99
 Информация в ЭВМ 157
 Исполнительный элемент 12

К

Качество управления 75
 Квантование непрерывных сигналов 100
 — — — по времени 100
 — — — по времени и уровню 101
 — — — по уровню 100
 Колебательность 33
 Корневые методы 76
 Корректирующие устройства 12
 — — параллельные 12, 81
 — — последовательные 12, 81
 Коррекция импульсных систем 114
 — линейных систем 81
 Косвенные оценки качества 76
 Коэффициентный метод построения ЛАЧХ 64
 Коэффициенты передаточной функции 52
 — — передачи звена 30
 — — усиления звена 30
 Критерии устойчивости 67
 — — алгебраические 67
 — — Гурвица 68
 — — Михайлова 69
 — — Найквиста 69
 — — Рауса 67
 — — частотные 69

Л

Линейные системы управления 40
 Линии переключения 91
 Логарифмическая амплитудно-частотная характеристика (ЛАЧХ) 57
 — — фазочастотная характеристика (ЛФЧХ) 57
 Логарифмический критерий устойчивости 72
 Локальный минимум целевой функции 240

М

Матричное представление модели 187
 Метод ветвей и границ 232
 — вещественных частотных характеристик 78
 — гармонического баланса 93
 — Гомори 230
 — градиента 136, 241
 — динамического программирования 127
 — Жордана — Гауса 191
 — критического пути 224

- математического программирования 180
- минимального элемента 216
- неявного перебора вариантов 233
- операторный 76
- отсекающих плоскостей 230
- Попова 97
- трапециевидных частотных характеристик 80, 114
- фазовых траекторий (портретов) 87
- штрафных функций 243
- Эрроу — Гурвица 247
- Микропроцессор 158
- МикроЭВМ 158
- Мини-ЭВМ 156
- Модели аналитические детерминированные 284
- математические 34
- математические сложных систем 39
- линейные 40, 179
- Модуляция амплитудно-импульсная 101
- временно-импульсная 101
- широтно-импульсная 101

Н

- Надежность АСУ 294, 296
- комплекса технических средств 296
- организационно-технологическая 297
- эксплуатационная 297
- Накопитель на магнитных барабанах 157
- — — дисках 157
- — — картах 157
- — — лентах 157
- Нелинейные системы 84
- Нелинейный элемент 85
- Номограмма построения фазовых характеристик 58

О

- Обеспечение АСУ информационное 282
- — лингвистическое 286
- — математическое 284
- — организационное 288
- — правовое 288
- — программное 285
- — техническое 282
- Обеспечение САПР информационное 268, 275
- — лингвистическое 268, 270
- — математическое 268, 271
- — методическое 268
- — организационное 268

- — программное 268, 273
- — техническое 268, 276
- Область устойчивости 72
- Обратная связь 10
- — гибкая 19
- — главная 11
- — жесткая 19
- — местная 12, 13
- — отрицательная 11
- — положительная 12
- Объект управления 10
- Оператор характеристический 53
- Определители Гурвица 68
- Оптимальные системы 116
- Оптимизаторы 131
- Особая точка 89
- Оценки качества управления 76
- — — косвенные 76
- — — прямые 76
- Ошибка регулирования в замкнутой системе 32
- — в разомкнутой системе 32
- среднеквадратичная 34
- статическая 32

П

- Передающая функция 41
- — аperiodического звена 41
- — дифференцирующих звеньев 47
- — замкнутой системы 51
- — интегрирующего звена 45
- — колебательного звена 43
- — усилительного звена 45
- Переменные состояния 37, 122
- Перерегулирование 34
- Переходная функция 41
- Плоскость фазовая 88
- Подсистемы АСУ 281
- Подсистемы САПР 267
- — обслуживающие 267
- — проектирующие 267
- Показатели динамических режимов САУ 33
- Портрет фазовый 89
- Порядок астатизма 57
- Постоянная времени звена 41
- Предельные циклы 90
- Преобразование Лапласа 35
- непрерывного сигнала 100
- структурных схем 50
- Преобразователь аналого-цифровой 170
- цифро-аналоговый 171
- Принцип максимума Понтрягина 123
- оптимальности Беллмана 127
- управления комбинированный 24
- — по возмущению 24
- — — отклонению 22
- Программирование математическое 39
- — динамическое 123
- — линейное 180

- — пелинейное 239
- — стохастическое 251
- — целочисленное 228
- Пространство фазовое 89
- Процесс проектирования 263
- Процессор 156

Р

- Разностные уравнения 102
- Разность второго порядка 103
 - первого порядка 103
- Регулируемая величина 10
- Регулятор 10
- Режимы работы систем управления 29
 - — — — динамические 29
 - — — — статические 29
- Релейные системы 100
- Релейный элемент 100
- Решетчатые функции 102

С

- Сепаратрисы 91
- Синтеза задача 35, 121
- Система автоматизированного проектирования 261
 - автоматизированного управления 14, 15
 - автоматического регулирования 12, 15
 - автоматического управления 12, 15
 - адаптивная 130
 - астатическая 28
 - динамическая 34
 - дискретная 99
 - линейная 40
 - нелинейная 84
 - оптимальная 116
 - самонастраивающаяся 146
 - самообучающаяся 146
 - статическая 28
- Соединения звеньев общий случай 50
 - — параллельное 49
 - — последовательное 49
 - — с перекрестными обратными связями 50
- Стабилизация САУ 33
- Статика системы 29
- Статическая характеристика 30
- Схема функциональная 12

Т

- Таблица коэффициентов эквивалентных АФХ 96
 - матриц альтернатив решений 292
 - типовых нелинейностей 85
- Теорема Куна — Таккера 245

- Технико-экономическое обоснование АСУ 288
- Техническое задание на разработку АСУ 288
- Типовая единичная трапеция 79
- Типовые нелинейности 85
- Типовые проектные решения АСУ 290

У

- Универсальность МО 271
- Управление 5
 - автоматизированное 14
 - автоматическое 12
 - адаптивное 130
 - астатическое 28
 - оптимальное 116
 - по замкнутому циклу 10
 - по разомкнутому циклу 10
 - ручное 10
- Уравнения Беллмана 129
 - линейные 48
 - нелинейные 84
 - статики 31
- Устойчивость 33
 - «в большом» 87
 - «в малом» 87
 - импульсных систем 109
 - линейных систем 65
- Устройство ввода 156
 - вывода 156
 - запоминающее 156
 - — внешнее 156
 - — оперативное 156
 - — сверхоперативное 157
 - поиска экстремума 134
 - приема цифровой информации 172
 - текущего времени 172
 - формирования входного сигнала 144

Х

- Характеристический оператор системы 53

Ц

- Цикл предельный 90
 - — неустойчивый 90
 - — устойчивый 90
- Цифровые ЭВМ 151
- Целевая функция 179

Ф

- Фазовые сдвиги 62
 - траектории 88
- Фильтрующие свойства элементов 93
- Функциональная схема системы 12
- Функциональные элементы САУ 12

Ч

Частота автоколебаний 93

— среза 72

— сопрягающая 58

Частотные критерии устойчивости 59

— характеристики систем 54

Э

Экстремальные системы управления
130

Электродвигатель как объект управления 16

Электромашина схема управления
17

Элемент исполнительный 12

— управляющий 12

— усилительный 12

— функциональный 12

Я

Языки программирования 270

— — высокого уровня 270

— — машинно-ориентированные 286

— — машинные 286

— — проблемно-ориентированные 286

Языки проектирования 270

— — внутренние 271

— — входные 270

— — выходные 271

— — сопровождения 271

ОСНОВНЫЕ АББРЕВИАТУРЫ

АБД	— автоматизированный банк данных		тизированные системы управления
АВМ	— аналоговая вычислительная машина	ИУ	— исполнительное устройство
АИС	— автоматизированная информационная система	ИЭ	— импульсный элемент
АСОУ	— автоматизированная система организационного управления	КУ	— корректирующее устройство
АСУП	— автоматизированная система управления предприятием	ЛАФЧХ	— логарифмическая амплитудно-фазовая частотная характеристика
АСУ ТП	— автоматизированная система управления технологическими процессами	ЛАЧХ	— логарифмическая амплитудно-частотная характеристика
АРМ	— автоматизированное рабочее место	ЛП	— линейное программирование
АФХ	— амплитудно-фазовая характеристика	ЛФЧХ	— логарифмическая фазо-частотная характеристика
АЦП	— аналого-цифровой преобразователь	ЛЧ	— линейная часть системы
АЧХ	— амплитудно-частотная характеристика	МКЭ	— метод конечных элементов
БД	— база данных	МО	— математическое обеспечение
ВЗУ	— внешнее запоминающее устройство	НМБ	— накопитель на магнитном барабане
ВИМ	— временно-импульсная модуляция	НМД	— накопитель на магнитном диске
ГАП	— гибкое автоматизированное производство	НМЛ	— накопитель на магнитной ленте
ГОС	— главная обратная связь	НЭ	— нелинейный элемент
ГПС	— гибкие производственные системы	ОВ	— обмотка возбуждения
ДС	— дискретные системы	ОПМО	— общее программно-математическое обеспечение
ИАСУ	— интегрированные автома-	ОР	— объект регулирования
		ОС	— обратная связь

ППП	— пакет прикладных программ
САПР	— система автоматизированного проектирования
САР	— система автоматического регулирования
САУ	— система автоматического управления
СПМО	— специальное программно-математическое обеспечение
ТЗ	— техническое задание
ТЛП	— технологическая линия проектирования

ТЭО	— технико-экономическое обоснование
УПТ	— усилитель постоянного тока
ЦАП	— цифро-аналоговый преобразователь
ЧИМ	— частотно-импульсная модуляция
ШИМ	— широтно-импульсная модуляция
ЭВМ	— электронно-вычислительная машина

1. Автоматизация проектирования строительных и технических объектов / Д. А. Аветисян, В. П. Игнатов, Г. Д. Фролов, Г. Я. Эпельцвейг — М.: Наука, 1986.— 135 с.
2. Автоматизированные системы управления предприятиями / Под ред. В. Н. Четверикова.— М.: Высш. шк., 1979.— 303 с.
3. Человеческий и оптимальный синтез на ЭВМ систем управления / Под ред. А. А. Воронова и И. А. Огурка.— М.: Наука, 1984.— 344 с.
4. Бесекерский В. А., Попов Е. П. Теория систем автоматического регулирования.— М.: Наука, 1975.— 767 с.
5. Воронов А. А., Титов В. К., Новогранов Б. Н. Основы теории автоматического регулирования и управления.— М.: Высш. шк., 1977.— 519 с.
6. Вычислительный комплекс ЛИРА / Сост.: А. С. Городецкий, М. Е. Винницкий.— К.: Реклама, 1984.— 16 с.
7. Гибкие автоматизированные производственные системы / Под ред. Л. С. Ямпольского.— К.: Техніка, 1985.— 280 с.
8. Голубничий Н. И., Селиванов В. Л. Основы моделирования систем автоматического регулирования.— К.: Вища шк. Головное изд-во, 1984.— 160 с.
9. Дегтярев Ю. И. Исследование операций.— М.: Высш. шк., 1986.— 320 с.
10. Заморин А. П. и др. Вычислительные машины, системы, комплексы: Справочник / А. П. Заморин, А. А. Мячев, Ю. П. Селиванов; Под ред. Б. Н. Наумова, В. В. Пржиялковского.— М.: Энергоатомиздат, 1985.— 264 с.
11. Куропаткин П. В. Оптимальные и адаптивные системы.— М.: Высш. шк., 1980.— 287 с.
12. Майника Э. Алгоритмы оптимизации на сетях и графах.— М.: Мир, 1981.— 323 с.
13. Маудер Дж., Элмаграби С. Исследование операций. Т. 1. Методологические основы и математические методы.— М.: Мир, 1981.— 712 с.
14. Михайлов В. С., Билецкий О. Б. Основы построения и проектирования автоматизированных систем управления в строительстве.— К.: Вища шк. Головное изд-во, 1984.— 392 с.
15. Михайлов В. С., Сокирко В. А. Синтез функциональной структуры АСУ // Механизация и автоматизация упр.— 1982.— № 3.— С. 27—31.
16. Михайлов В. С., Товченко В. И. Автоматизация разработки организационно-технологических вариантов возведения зданий // Изв. вузов. Строительство и архитектура.— 1982.— № 2.— С. 87—90.
17. Михайлов В. С., Товченко В. И., Никифорова Л. В. Имитационные модели оперативного управления строительным производством в режиме диалога // Там же.— 1983.— № 9.— С. 74—77.
18. Морозов В. П., Дымарский Я. С. Элементы теории управления ГАП: Математическое обеспечение.— Л.: Машиностроение. Ленингр. отд-ние, 1984.— 333 с.
19. Норенков И. П. Введение в автоматизированное проектирование технических устройств и систем.— 2-е изд.— М.: Высш. шк., 1986.— 304 с.
20. Норенков И. П., Маничев В. Б. Системы автоматизированного проектирования электронной и вычислительной аппаратуры.— М.: Высш. шк., 1983.— 272 с.
21. Основы автоматизации управления производством / Под ред. И. М. Макарова.— М.: Высш. шк., 1983.— 504 с.

22. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность.— М.: Мир, 1985.— 512 с.
23. Петренко А. И. Основы автоматизации проектирования.— К.: Техніка, 1982.— 293 с.
24. Петренко А. И., Семенов О. И. Основы построения систем автоматизированного проектирования.— К.: Вища шк. Головное изд-во, 1985.— 294 с.
25. Прангшвили И. В. Микропроцессоры и локальные сети микроЭВМ в распределенных системах управления.— М.: Энергоатомиздат, 1985.— 272 с.
26. Системы автоматизированного проектирования: Типовые элементы, методы и процессы.— М.: Изд-во стандартов, 1985.— 180 с.
27. Солодовников В. В., Плотников В. Н., Яковлев А. В. Основы теории и элементы систем автоматизированного регулирования.— М.: Машиностроение, 1985.— 536 с.
28. Таха Х. Введение и исследование операций.— М.: Мир, 1985.— Кн. 1.— 479 с.
29. Теория автоматического управления / Под ред. А. А. Воронова.— М.: Высш. шк.— Ч. 1.— 1977.— 303 с.; Ч. 2.— 1977.— 288 с.
30. Филлипс Д., Гарсия-Диас А. Методы анализа сетей.— М.: Мир, 1984.— 496 с.
31. Фритч В. Применение микропроцессоров в системах управления: Пер. с нем.— М.: Мир, 1984.— 464 с.
32. Цифровые электромеханические системы / В. Г. Каган, Ю. Д. Бери, Б. И. Акимов и др.— М.: Энергоатомиздат, 1985.— 208 с.
33. Чаки Ф. Современная теория управления.— М.: Мир, 1975.— 424 с.
34. ЭВМ в проектировании и производстве // Под общ. ред. Г. В. Орловского.— Л.: Машиностроение. Ленингр. отд-ние, 1983.— 296 с.
35. Энциклопедия кибернетики: В 2 т.— К.: Гл. ред. Укр. сов. энцикл., 1975.— Т. 1.— 606 с.; Т. 2.— 619 с.

От автора	3
Краткий обзор развития и современного состояния теории управления	5
Глава 1. Основные определения и понятия теории управления	10
1.1. Объекты и процессы управления. Виды управления	10
1.2. Примеры элементов и систем управления	16
1.3. Принципы и законы управления (регулирования)	22
1.4. Режимы работы систем управления	29
1.5. Постановка и методы решения задач управления	34
Глава 2. Линейные динамические системы управления	40
2.1. Математические модели звеньев линейных динамических систем	40
2.2. Математические модели линейных динамических систем	48
2.3. Частотные характеристики линейных динамических систем	54
2.4. Устойчивость линейных динамических систем	65
2.5. Качество управления в линейных динамических системах	75
2.6. Коррекция линейных динамических систем	81
Глава 3. Нелинейные динамические системы управления	84
3.1. Определение и особенности нелинейных динамических систем управления	84
3.2. Исследование нелинейных динамических систем методом фазовых траекторий	87
3.3. Метод гармонического баланса	93
Глава 4. Дискретные системы автоматического управления	99
4.1. Особенности дискретного управления. Квантование непрерывных величин. Модуляция импульсов	99
4.2. Математическое описание систем дискретного управления	102
4.3. Уравнение и передаточные функции импульсных систем	106
4.4. Устойчивость импульсных систем	109
4.5. Качество управления в импульсных системах	114
Глава 5. Оптимальное управление динамическими системами	116
5.1. Основные понятия систем оптимального управления	116
5.2. Задачи синтеза оптимальных систем и математические модели объектов управления	121
5.3. Методы решения задач оптимального управления динамическими системами	123
5.4. Экстремальные автоматические системы управления	130
5.5. Переходные процессы в экстремальных автоматических системах управления	137
5.6. Устойчивость экстремальных систем. Улучшение качества управления в экстремальных системах	140
5.7. Самонастраивающиеся и самообучающиеся динамические системы оптимального управления	146

Глава 6. Электронно-вычислительные машины в теории и практике управления динамическими системами	151
6.1. Типы электронно-вычислительных машин	151
6.2. Исследование динамических систем управления с использованием аналоговых вычислительных машин	159
6.3. Исследование динамических систем управления с использованием цифровых электронно-вычислительных машин	162
6.4. Управление динамическими системами на базе ЭВМ	170
Глава 7. Математические методы решения задач управления сложными системами	179
7.1. Общая характеристика. Основы метода линейного программирования	179
7.2. Оптимизация на сетях	210
7.3. Целочисленное программирование и комбинаторная оптимизация	228
7.4. Нелинейное программирование	239
7.5. Стохастическое программирование	251
Глава 8. Системы автоматизированного проектирования	259
8.1. Общая характеристика систем автоматизированного проектирования	259
8.2. Схема процесса проектирования и элементный состав САПР	263
8.3. Структура систем автоматизированного проектирования. Стадии создания САПР	266
8.4. Основные компоненты систем автоматизированного проектирования	270
8.5. Примеры систем автоматизированного проектирования	277
Глава 9. Автоматизированные системы управления	279
9.1. Общая характеристика и классификация автоматизированных систем управления	279
9.2. Обеспечивающая часть АСУ	282
9.3. Основы разработки АСУ	288
9.4. Критерии принятия решений в условиях неопределенности	292
9.5. Гибкие автоматизированные производства	300
<i>Предметный указатель</i>	300
<i>Основные аббревиатуры</i>	306
<i>Список использованной литературы</i>	307

Учебное издание

Михайлов Виталий Степанович

ТЕОРИЯ УПРАВЛЕНИЯ

Переплет художника *Г. М. Балюна*
Художественный редактор *С. П. Духленко*
Технический редактор *О. В. Козлитина*
Корректор *С. Г. Чиркина*

ИБ № 12473

Сдано в набор 09.11.87. Подписано в печать 15.11.88.
БФ 03680. Формат 60×90¹/₁₆. Бум. тип. № 2. Гарнитура
литературная. Высокая печать. Усл. печ. л. 19,5. Усл.
кр.-отт. 19,5. Уч.-изд. л. 21,14. Тираж 3000 экз. Изд.
№ 8060. Зак. 260. Цена 1 р. 10 к

Главное издательство издательского объединения «Выща
школа», 252054, Киев-54, ул. Гоголевская, 7

Белоцерковская книжная фабрика, 256400, Белая Цер-
ковь, ул. Карла Маркса, 4

Головное издательство
издательского объединения «Выща школа»
готовит к выпуску в свет новую книгу:

Товченко В. И., Михайлов В. С. Модели и алгоритмы управления строительным производством.— Монография.— К.: Выща шк. Головное изд-во.—10 л., ил.— Яз. рус.—2 р.

В монографии приведены модели и методы исследования сложных строительных систем управления, информационные характеристики, необходимые для моделирования процессов возведения зданий и сооружений, теоретико-множественное представление исходной информации в виде неориентированного графа. Даны способы контроля правильности и корректировки данных в информационной модели.

Большое внимание уделено моделям автоматизации проектирования схем возведения зданий и сооружений, многовариантным моделям организационно-технологических схем строительства.

Освещены вопросы моделирования оперативного управления строительным производством с широким применением интерактивных режимов, вопросы построения систем управления строительной организацией на базе локальных мини- и микроЭВМ с коллективным доступом.

Книга представляет интерес для инженерно-технических, научных работников, аспирантов, занимающихся разработкой и эксплуатацией систем управления на базе современных средств вычислительной техники, а также может быть полезна руководителям строительных организаций и студентам старших курсов технических вузов.

Авторы: В. И. Товченко — канд. техн. наук, доц. Киевского инженерно-строительного института;

В. С. Михайлов — д-р техн. наук, проф. Киевского инженерно-строительного института